

INNEHÅLLSFÖRTECKNING

I	MULTIPEL KORRELATION	
	1. Introduktion.....	1
	2. Två prediktorer uttryckta i standardpoäng...	2
	3. Grafisk representation.....	6
	4. Två prediktorer uttryckta i råpoäng.....	7
	5. Fler än två prediktorer.....	10
	6. Signifikanstestningar.....	12
II	PARTIELL KORRELATION	
	1. Introduktion.....	14
	2. En konstanthållen variabel.....	14
	3. Två konstanthållna variabler.....	17
III	SEMIPARTIELL KORRELATION	
	1. Introduktion.....	18
	2. Beräkning av semipartiell korrelation.....	18
	3. Förhållandet mellan semipartiell och multipel korrelation.....	20
	4. Alternativa beräkningssätt för partiell och semipartiell korrelation.....	23
IV	KANONISK KORRELATION	
	1. Introduktion.....	26
	2. Exemplifiering.....	27
V	AVSLUTANDE SYNPUNKTER.....	30
	ÖVNINGSUPPGIFTER.....	34
	TABELLER.....	39

Kapitel I

MULTIPEL KORRELATION

1. Introduktion

De vanligaste förekommande sambandsmått, t.ex. produktmoment-korrelation, rangkorrelation och biserial korrelation, används för att uttrycka ett samband mellan två variabler. Vi skall i detta kompendium redogöra för några sambandsmått som används när man har tillgång till data från tre eller fler variabler. I sådana fall kan man göra s.k. multivariata analyser och få fram skilda typer av korrelationskoefficienter, koefficienter som på olika sätt belyser sambandsstrukturen mellan variablerna. En redogörelse kommer att ges för Multipel, Partiell, Semipartiell och Kanonisk korrelation. Dessa fyra torde vara några av de mest använda multivariata sambandsmått, men utgör endast ett mindre antal av samtliga förekommande. Andra multivariata metoder presenteras i den litteratur som refereras i kapitel V.

En multipel korrelationskoefficient upplyser oss om sambandet mellan en variabel och en kombination av ett antal andra variabler. Närmare bestämt ger den oss ett mått på hur väl värdet i en kriterievariabel kan förutsägas med utgångspunkt från en grupp prediktionsvariabler, vilka vägs samman på ett lämpligt sätt.

Filosofin bakom den multipla korrelationen respektive regressionen är följande: För att kunna klara av en uppgift bra krävs i allmänhet att man besitter ett flertal egenskaper. För att bli en bra lärare fordras inte endast ämneskunskaper, utan även undervisningsskicklighet, objektivitet, förmåga att umgås med elever etc. För att bli en skicklig kirurg, poliskommisarie, skådespelare eller bilmekaniker krävs något annorlunda uppsättningar av egenskaper. I tekniska termer innebär detta, att man bör ha relativt höga värden i ett flertal prediktionsvariabler, för att nå ett högt värde i en bestämd kriterievariabel.

Av resonemanget följer också, att man i en urvalssituation kan göra en bättre selektion, om man har tillgång till flera prediktorer, vilka mäter olika egenskaper hos de sökande - egenskaper som alla är av vikt för den kommande befattningen.

2. Två prediktorer uttryckta i standardpoäng

När man beräknar en multipel korrelation tilldelas prediktorerna olika vikter. Väsentliga egenskaper, dvs. prediktorer som korrelerar högt med kriterievariabeln får större vikter, mindre väsentliga egenskaper tilldelas lägre vikter. Dessa vikter kallas β -vikter.

Om man har två prediktorer som båda är uttryckta i standardpoäng, erhåller man individens väntade värde i kriterievariabeln, genom att helt enkelt multiplicera de funna värdena i x-variablerna med respektive β -vikt, varefter man adderar produkten.

$$Z^1_Y = \beta_1 \cdot Z_{x_1} + \beta_2 \cdot Z_{x_2}$$

Exempel: Vid en nystartad institution för telegrafistutbildning antar man första året samtliga 80 sökande. Innan utbildningen startar får var och en genomgå ett test som mäter akustisk diskrimination och ett test som mäter reaktionshastighet. Efter 10 månaders utbildning erhåller de studerande sina avgångsbetyg. Korrelationerna mellan betyg och respektive test samt mellan testen inbördes framgår nedan:

Avgångsbetyg (y) - Akustisk kapacitet (x_1) : 0.60

Avgångsbetyg (y) - Reaktionshastighet (x_2) : 0.50

Akustisk kapacitet (x_1) - Reaktionshastighet (x_2) : 0.20

Vid nästa ansökningstillfälle söker 120 personer, men man beslutar sig för att endast anta 50. Vidare beslutar man sig för att använda de båda testen som prediktorer. För att få bästa möjliga urval, dvs. ett så högt samband som möjligt mellan testresultat och avgångsbetyg, använder man tekniken med multipel regression. Först måste man då beräkna β -vikterna.

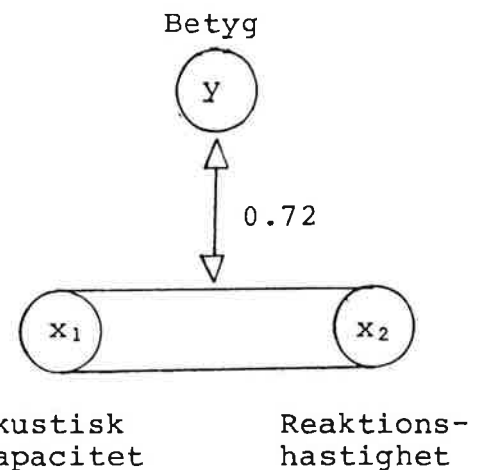
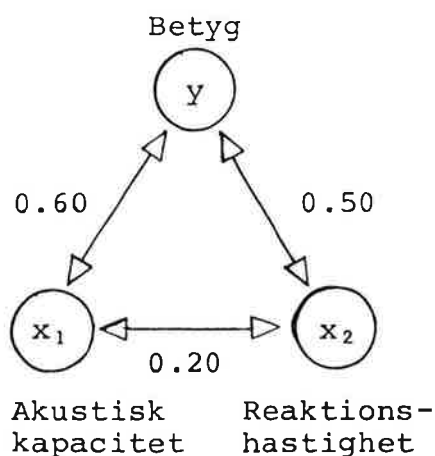
$$\beta_1 = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0.60 - 0.50 \cdot 0.20}{1 - 0.20^2} = 0.52$$

$$\beta_2 = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0.50 - 0.60 \cdot 0.20}{1 - 0.20^2} = 0.40$$

Den multipla korrelationen erhålles sedan genom formeln:

$$R = \sqrt{\beta_1 \cdot r_{yx_1} + \beta_2 \cdot r_{yx_2}} = \sqrt{0.52 \cdot 0.60 + 0.40 \cdot 0.50} = 0.72$$

Resultatet innebär att den multipla korrelationen mellan betyg och de båda testen uppgår till 0.72, vilket kan jämföras med de enskilda prediktorernas samband med kriteriet (0.60 respektive 0.50).



För att erhålla maximalt samband - i detta fall $R = 0.72$ - måste dock prediktorerna ingå med respektive β -vikter. Detta innebär att varje individs testpoäng i det akustiska testet multipliceras med 0.52 och varje poäng i reaktionstestet med 0.40. Innan vägningen sker måste dock testresultaten uttryckas i z-skalan ($\bar{X} = 0$, $s = 1$).

I tabllån nedan ges råpoängen, z-poängen och den vägda z-poängen för fyra individer i de båda testen.

Akustisk diskrimination : $\bar{X} = 120$, $s = 20$

<u>Ind</u>	<u>Råpoäng</u>	<u>z-poäng</u>	<u>vägd z-poäng</u>
A	140	1.0	0.52
B	148	1.4	0.73
C	120	0.0	0.00
D	128	0.4	0.21

Reaktionshastighet : $\bar{X} = 150$, $s = 30$

<u>Ind</u>	<u>Råpoäng</u>	<u>z-poäng</u>	<u>vägd z-poäng</u>
A	195	1.5	0.60
B	180	1.0	0.40
C	165	0.5	0.20
D	150	0.0	0.00

Vilken person har bäst förutsättningar att lyckas med telegrafistutbildningen?

Summan av A:s vägda z-poäng blir:

$$\beta_1 \cdot z_{x_1} + \beta_2 \cdot z_{x_2} = 0.52 \cdot 1.0 + 0.40 \cdot 1.5 = 0.52 + 0.60 = 1.12$$

Poängen 1.12 är individens väntade värde i kriterievariabeln - den bästa uppskattning som man kan göra med utgångspunkt från hans resultat i de båda prediktionsvariablerna.

Om avgångsbetyget är uttryckt i en skala med medeltalet 50 och spridningen 10 blir A:s väntade värde:

$$1.12 = \frac{X - 50}{10}$$

$$X = 1.12 \cdot 10 + 50 = 11.2 + 50 = 61.2$$

Hur högt hade sambandet blivit om vi ej vägt prediktorerna med respektive β -vikter? Detta går relativt lätt att beräkna med

hjälp av följande korrelationsmatris, där vi angett variabler-
nas interkorrelationer samt satt ett 1 i diagonalen.

	y	x ₁	x ₂
y	1.0	0.6	0.5
x ₁	0.6	1.0	0.2
x ₂	0.5	0.2	1.0

A	C
C	B

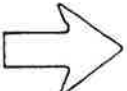
$$r_{z_Y(z_{x_1}+z_{x_2})} = \frac{C}{\sqrt{AB}} = \frac{0.6 + 0.5}{\sqrt{1.0(1.0+1.0+0.2+0.2)}} = \frac{1.1}{\sqrt{2.4}} = \frac{1.1}{1.55} =$$

$$= \underline{\underline{0.71}}$$

I detta fall ingår prediktorerna med lika vikt, ty spridningarna är lika ($s_{x_1} = s_{x_2} = 1$). Vid direkt sammanläggning bestämmer nämligen spridningarna de enskilda x-variablernas betydelse. Ju större spridning i x-variabeln, desto större vikt får den vid sammanläggningen.

I vårt exempel är spridningen i råpoäng i den första prediktorn 20 och i den andra 30. Om råpoängen hade adderats direkt hade vi fått följande samband.

		20	30
	y	x ₁	x ₂
y	1.0	0.6	0.5
20 x ₁	0.6	1.0	0.2
30 x ₂	0.5	0.2	1.0



A	C
1.0	12 15
C	B
12	400 120
15	120 900

I detta fall måste vi multiplicera med de aktuella spridningarna på det sätt som framgår ovan, varefter vi erhåller:

$$r_{Y(x_1+x_2)} = \frac{C}{\sqrt{AB}} = \frac{12 + 15}{\sqrt{1.0 (400+120+120+900)}} = \frac{27}{\sqrt{1540}} = \frac{27}{39.3} = \underline{\underline{0.69}}$$

Om vi i stället multiplicerar med de framräknade β -vikterna, får vi:

		0.52	0.40
	1.0	0.6	0.5
0.52	0.6	1.0	0.2
0.40	0.5	0.2	1.0

➔

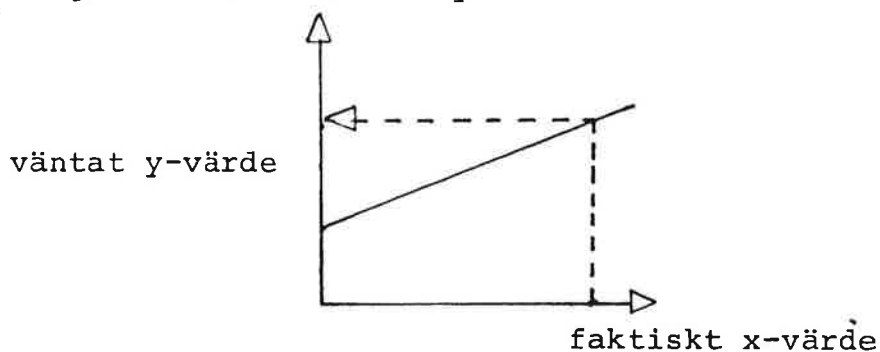
A	C	
1.0	0.31	0.20
C	B	
0.31	0.27	0.04
0.20	0.04	0.16

I detta - och endast i detta - fall blir
 $C = B = 0.51$ och sambandet maximeras

$$R = \frac{C}{\sqrt{AB}} = \frac{0.51}{\sqrt{0.51}} = \underline{\underline{0.72}}$$

3. Grafisk representation

Om man vill skatta en individs värde i en kriterievariabel utifrån hans värde i en prediktionsvariabel, kan man göra detta grafiskt, genom att lägga in den aktuella regressionslinjen i ett koordinatsystem.



Även när man har två prediktorer kan man åskådliggöra förhållandet grafiskt. I detta fall måste man bygga upp en tredimensionell figur och regressionslinjen utbyts mot ett regressionsplan.

I följande figur har vi lagt in kriterievariabeln (y) längs en vertikal axel. Vinkelrät mot denna såväl som mot varandra ligger de båda prediktionsvariablerna (x_1 och x_2). Vidare har vi markerat A:s värden i x_1 och x_2 . Utifrån dessa punkter drar vi linjer parallellt med respektive x -axel, tills vi når linjernas skärningspunkt (E). Härifrån drar vi en lodrät linje upp till ytan FGHI, det tidigare nämnda regressionsplanet. Planet stöter vi på vid punkten J, varefter vi avläser höjden EJ utefter y -axeln. Härigenom får vi veta A:s väntade värde i kriterievariabeln (1.12).

Regressionsplanets lutning bestäms av lutningarna hos linjerna FI och FG. Lutningen hos dessa linjer bestäms i sin tur av β_1 respektive β_2 , vilka alltså inte endast är z -poängens vikt-koefficienter utan även respektive linjers regressionskoefficienter.

4. Två prediktorer uttryckta i råpoäng

I vårt tidigare exempel transformerade vi först råpoängen till z -poäng för att få fram de väntade värdena i y -variabeln. Sedan förvandlade vi åter z -poängen till råpoäng för att få kunskap om individens väntade betyg. I och för sig är det ej nödvändigt att gå denna omväg via transformationer, utan man kan direkt skatta den väntade råpoängen.

Om man har en prediktionsvariabel och denna är uttryckt i z -poängen, blir det väntade värdet i kriterievariabeln:

$$z_y^1 = \beta \cdot z_x \quad \text{där } \beta = r_{xy}$$

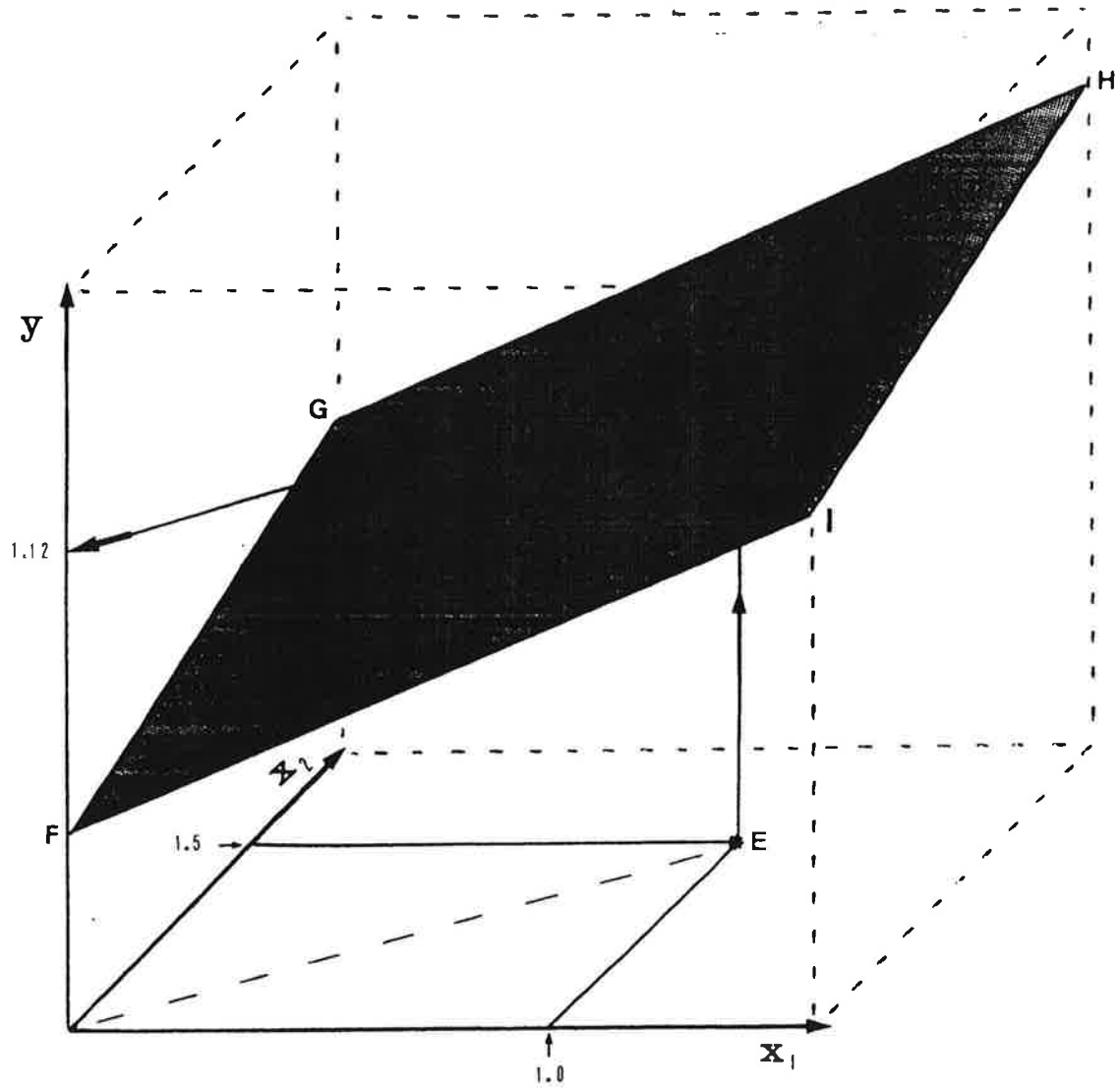
Om man i stället arbetar med råpoäng, blir det väntade y -värdet:

$$Y^1 = a + bX$$

där $a = \bar{Y} - b\bar{X}$

och $b = \beta \cdot \frac{s_y}{s_x}$

Grafisk representation av multipel regression



Exempel

$$\beta = 0.60, \quad z_x = 1.0; \quad z^1_y = 0.60 \cdot 1.0 = \underline{\underline{0.6}}$$

$$\bar{Y} = 50, \quad s_y = 10, \quad \bar{X} = 120, \quad s_x = 20, \quad X = 140$$

$$b = 0.60 \cdot \frac{10}{20} = 0.30$$

$$a = 50 - 0.30 \cdot 120 = 50 - 36 = 14$$

$$y^1 = 14 + 0.30 \cdot 140 = 14 + 42 = \underline{\underline{56}}$$

dvs. 0.6 spridningsenheter över medeltalet i y-variabeln.

Har man två prediktions- och en kriterievariabel, vilka samtliga är uttryckta i z-poäng, blir som vi tidigare sett:

$$z^1_y = \beta_1 \cdot z_{x_1} + \beta_2 \cdot z_{x_2}$$

Är däremot samtliga variabler uttryckta i råpoäng blir det väntade värdet i kriterievariabeln:

$$y^1 = a + b_1 X_1 + b_2 X_2$$

där $a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2$

$$b_1 = \beta_1 \frac{s_y}{s_{x_1}}$$

$$b_2 = \beta_2 \frac{s_y}{s_{x_2}}$$

Exempel

$$\beta_1 = 0.52, \quad \beta_2 = 0.40, \quad z_{x_1} = 1.0, \quad z_{x_2} = 1.5$$

$$z^1_y = 0.52 \cdot 1.0 + 0.40 \cdot 1.5 = \underline{\underline{1.12}}$$

$$\bar{Y} = 50, \quad s_Y = 10, \quad \bar{X}_1 = 120, \quad s_{X_1} = 20$$

$$\bar{X}_2 = 150, \quad s_{X_2} = 30, \quad X_1 = 140, \quad X_2 = 195$$

$$b_1 = 0.52 \cdot \frac{10}{20} = 0.260$$

$$b_2 = 0.40 \cdot \frac{10}{30} = 0.133$$

$$a = 50 - 0.260 \cdot 120 - 0.133 \cdot 150 = -1.15$$

$$Y^1 = -1.15 + 0.260 \cdot 140 + 0.133 \cdot 195 = \underline{\underline{61.20}}$$

Siffermaterialet här är hämtat från det tidigare exemplet och som framgår blir slutresultatet (individ A:s betyg) detsamma.

5. Fler än två prediktorer

Hittills har vi endast använt oss av två prediktionsvariabler, men man kan också utnyttja tre, fyra eller hur många prediktorer som helst för att beräkna en multipel korrelation. I princip bör ju också prognosen bli säkrare, desto fler prediktorer man har tillgång till.

Har man tre prediktorer erhålles den multipla korrelationen genom formeln:

$$R = \sqrt{\beta_1 \cdot r_{YX_1} + \beta_2 \cdot r_{YX_2} + \beta_3 \cdot r_{YX_3}}$$

Individens väntade z-värde blir:

$$z^1_Y = \beta_1 z_{X_1} + \beta_2 \cdot z_{X_2} + \beta_3 \cdot z_{X_3}$$

Uttryckt i råpoäng blir det väntade värdet:

$$Y^1 = a + b_1 X_1 + b_2 X_2 + b_3 X_3$$

Exempel Vid ett idrottsinstitut försöker man predicera ungdomars framtida höjdhoppsförmåga genom att använda sig av ett snabbhetsprov (x_1), en testning av muskelstyrkan (x_2) samt en mätning av benlängden (x_3). Genom en validitetsundersökning har man fått följande värden:

$$r_{yx_1} = 0.60, \quad r_{yx_2} = 0.50, \quad r_{yx_3} = 0.30$$

$$\beta_1 = 0.50 \quad \beta_2 = 0.40 \quad \beta_3 = 0.20$$

Den multipla korrelationen blir då:

$$R = \sqrt{0.50 \cdot 0.60 + 0.40 \cdot 0.50 + 0.20 \cdot 0.30} = \sqrt{0.56} = \underline{\underline{0.75}}$$

En yngling har z-värdet +1.0 i samtliga tre prediktorer. Vilken blir hans väntade z-poäng i kriterievariabeln? Vilket höjdhoppsresultat motsvarar i sin tur denna z-poäng, om medeltalet i kriterievariabeln är 150 cm med en spridning av 30 cm?

När man tar steget från två till tre prediktorer stöter man emellertid på en del svårigheter. En sådan svårighet är att det är ganska mödosamt att få fram värdena på β -vikterna. När man endast har två prediktorer är ju detta enkelt, men vid tre eller fler prediktorer blir det rena räknearbetet så omfattande att man i allmänhet överlämnar detta arbete åt datamaskinen.

En annan svårighet är att det är mycket svårt att använda sig av grafisk representation, eftersom man i princip behöver ha tillgång till en figur med fler än tre dimensioner.

Ytterligare en komplikation är att det oftast är svårt att finna en tredje - för att inte tala om en fjärde och en femte - prediktor som ger ett betydelsefullt tillskott till prediktionen. Det gäller ju att finna prediktorer som korrelerar högt med kriteriet, men som samtidigt är lågt korrelerade med redan utnyttjade prediktorer.

Uppgift. Kommentera det sista avsnittet och ge en kort redogörelse för den antydda svårigheten.

6. Signifikanstestningar

En multipel korrelationskoefficient signifikanstestas med följande formel:

$$F = \frac{R^2/k}{(1-R^2)/(N-k-1)}$$

där N = antalet individer eller observationer
 k = antalet prediktorer
 antalet frihetsgrader är k ; $N-k-1$

I vårt tidigare exempel med telegrafistutbildningen var $R = 0.72$, $k = 2$ och $N = 80$, vilket ger

$$F = \frac{0.72^2/2}{(1-0.72^2)/80-2-1} = \frac{0.52/2}{0.48/77} = \frac{0.26}{0.0062} = 41.94$$

Kritiskt värde ($df\ 2; 77$) = 3.11 (5 %)
 4.88 (1 %)

Resultatet av F-testningen innebär att den funna multipla korrelationen är starkt signifikant.

En annan fråga i detta sammanhanget är om ytterligare en prediktor ger en signifikant ökning av den multipla korrelationen. I nedanstående exempel undersöker vi om en tredje prediktor ger något signifikant bidrag, men den givna formeln kan användas generellt.

I exemplet ovan fann vi en multipel korrelation på 0.72, då vi använde de två prediktorerna akustisk diskrimination och reaktionshastighet. Antag att vi väljer ytterligare en prediktor, t.ex. induktiv begåvning och finner att den multipla korrelationen ökar till 0.76. Är denna ökning signifikant? Även i detta fall ger oss en F-testning svar.

$$F = \frac{(R^2_{Y.123} - R^2_{Y.12})/(k_1 - k_2)}{(1 - R^2_{Y.123})/(N - k_1 - 1)}$$

där $R_{Y.123}$ = den multipla korrelationen mellan y och tre prediktorer

$R_{Y.12}$ = den multipla korrelationen mellan y och två prediktorer

k_1 = det större antalet prediktorer

k_2 = det mindre antalet prediktorer

N = antalet observationer

$$F = \frac{(0.76^2 - 0.72^2)/(3 - 2)}{(1 - 0.76^2)/(80-3-1)} = \frac{(0.58 - 0.52)/1}{0.42/76} = \frac{0.06}{0.0055} = \underline{\underline{10.91}}$$

Kritiskt värde (df 1; 76) = 3.96 (5 %)

6.96 (1 %)

Anser Du att vi bör använda oss även av ett induktivt begåvnings-test vid urvalet till telegrafistutbildningen?

Kapitel II

PARTIELL KORRELATION

1. Introduktion

Inom pedagogisk forskning är det ofta omöjligt att tillämpa experimentell kontroll med t.ex. slumpvis utvalda kontroll- och experimentgrupper. I stället får man ta till statistisk kontroll, vilket innebär att man med statistisk metodik försöker eliminera icke önskvärda variationskällor. En form av statistisk kontroll är kovariansanalys. En annan är tekniken med partiella korrelationer.

Den partiella korrelationen används då man vill studera sambandet mellan två variabler samtidigt som man konstanthåller en tredje. Den variabel som konstanthålles samvarierar med de båda andra, vilket innebär att den påverkar sambandet mellan dessa. Genom att beräkna en partialkorrelation kan man eliminera inflytandet från "störningsvariabeln".

Med partiell korrelationsteknik kan man också konstanthålla fler än en störningsvariabel, när man studerar ett aktuellt samband. Beräkningsarbetet blir då emellertid ganska tidsödande.

2. En konstanthållen variabel

För att undersöka sambandet mellan längd och vikt bland barn i förskoleåldern, väljer man ut 500 barn i åldern 3 till 6 år. Man finner att sambandet uttryckt i en produktmomentkorrelation uppgår till 0.70. Denna koefficient ger dock en överskattning av sambandet, ty både längd och vikt är beroende av åldern. Man beräknar därför den partiella korrelationen mellan längd och vikt med konstanthållande av åldern. Härvid måste man först räkna ut korrelationerna mellan ålder och längd samt mellan ålder och vikt. Dessa korrelationer uppgår till 0.75 respektive 0.65.

Partialkorrelationen beräknas sedan med följande formel:

$$r_{12 \cdot 3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

$r_{12 \cdot 3}$ = partialkorrelationen mellan variabel 1 och 2, när variabel 3 konstanthålles

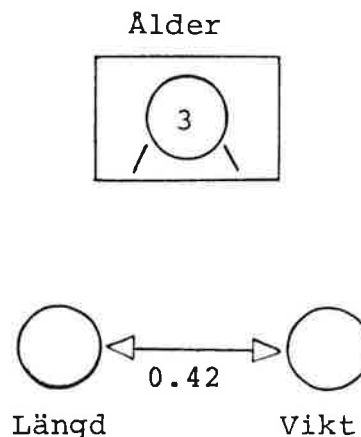
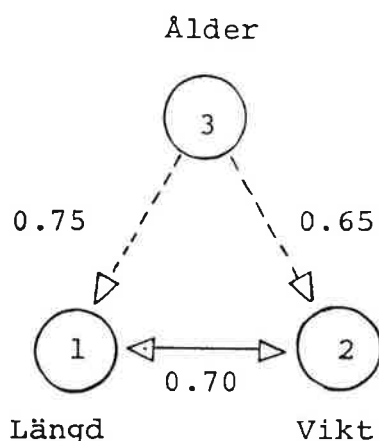
r_{12} = korrelationen mellan variabel 1 och 2

r_{13} = korrelationen mellan variabel 1 och 3

r_{23} = korrelationen mellan variabel 2 och 3

$$r_{12 \cdot 3} = \frac{0.70 - 0.75 \cdot 0.65}{\sqrt{(1-0.75^2)(1-0.65^2)}} = 0.42$$

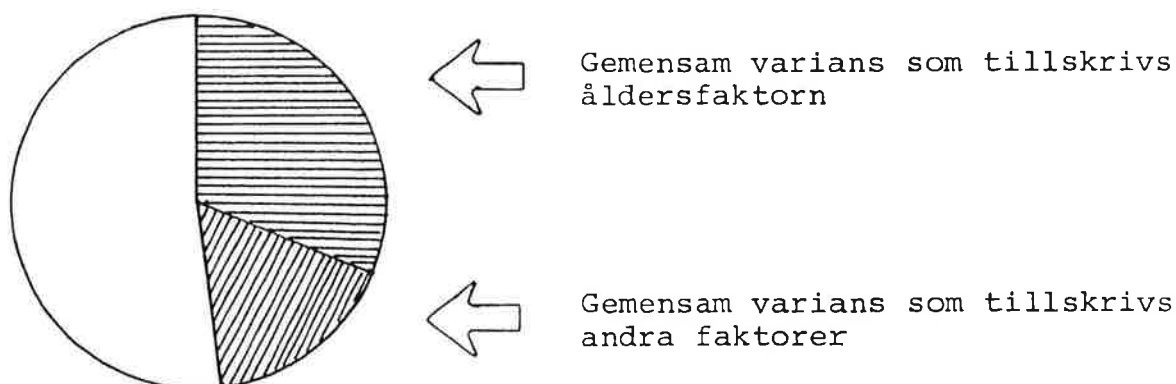
När åldern konstanthålles sjunker sålunda sambandet mellan längd och vikt från 0.70 till 0.42, vilket vi försökt åskådliggöra i nedanstående figurer:



Vilka upplysningar ger oss nu denna partiella korrelationskoefficient. Hur skall vi tolka den? Möjligen blir vi något klokare om vi tar steget via "förklarad varians".

Den ursprungliga korrelationen var 0.70. Kvadrerar vi denna koefficient och multiplicerar med 100, får vi sambandet uttryckt i form av gemensam varians - i detta fall 49 %.

Den partiella korrelationen sjunker till 0.42, vilket motsvarar 18 % gemensam varians. När vi konstanthåller åldern sjunker således den gemensamma variansen från 49 % till 18 % dvs. en reduktion med 31 %. 31/49 eller cirka 2/3 av den gemensamma variansen i vikt och längd kan därför tillskrivas åldersfaktorn, medan 1/3 får tillskrivas andra faktorer. Förhållandet åskådliggörs i nedanstående cirkeldiagram.



Ett annat sätt att se på den partiella korrelationen är följande:

I den exemplifierade undersökningen har man blandat barn från olika åldersgrupper och sedan konstanthållit åldern på statistisk väg. Om man i stället gjort separata korrelationsberäkningar inom varje åldersgrupp och sedan beräknat medeltalet av samtliga korrelationer, så hade man fått just koefficienten 0.42.

En partialkorrelation signifikant testas med formeln:

$$t = \frac{r_{12 \cdot 3}}{\sqrt{\frac{1 - r_{12 \cdot 3}^2}{N - 3}}}$$

Antal frihetsgrader är $N - 3$.

I vårt fall:

$$t = \frac{0.42}{\sqrt{\frac{1 - 0.42^2}{500 - 3}}} = 10.5$$

Uppgift. Kommentera utfallet av signifikanstestningen.

3. Två konstanthållna variabler

Som tidigare påpekats så är det tekniskt möjligt att konstanthålla fler än en variabel, när man beräknar en partiell korrelation. Detta förekommer relativt sällan, men vi skall dock skissera ett exempel.

Antag att man är intresserad av sambandet mellan resultat i längdhopp (variabel 1) och antal träningstimmar (variabel 2). Av någon anledning vill man dock ha reda på hur högt detta samband blir om man konstanthåller benlängd (variabel 3) och vikt (variabel 4).

Partialkorrelationen mellan variabel 1 och 2 när variabel 3 och 4 konstanthålles skrives $r_{12.34}$. den erhålles genom följande formel:

$$r_{12.34} = \frac{r_{12.4} - r_{13.4} r_{23.4}}{\sqrt{(1 - r_{13.4}^2)(1 - r_{23.4}^2)}}$$

Innan man kan använda sig av formeln måste man räkna ut de tre partiella korrelationerna $r_{12.4}$, $r_{13.4}$ och $r_{23.4}$. Dessa kallas för partiella korrelationer av första ordningen, till skillnad från $r_{12.34}$ som är en partiell korrelation av andra ordningen.

Kapitel III

SEMIPARTIELL KORRELATION

1. Introduktion

Vi skall nu övergå till något som kallas för semipartiella korrelationer. Anledningen till att vi redogör för denna typ av korrelation är att den utgör en förbindelselänk mellan den multipla och den partiella korrelationen. Den kan härigenom hjälpa till att ytterligare belysa likheter och skillnader mellan dessa båda sambandsmått. Det främsta skälet till att vi tar upp den semipartiella korrelationen är dock, att den torde öka förståelsen för den multipla korrelationens och regressionens uppbyggnad.

2. Beräkning av semipartiell korrelation

Vid en teknisk skola räknar man ut sambandet mellan betyg efter avslutad utbildning (variabel 1) och resultat på ett prov som ges innan utbildningen startar (variabel 2). Detta samband uppgår till 0.50 (r_{12}). Vidare är man intresserad av att få veta vad som händer när elevernas begåvningsmässiga förutsättningar (variabel 3) konstanthålles, varför man också beräknar sambandet mellan slutbetyg och begåvning (r_{13}) respektive provresultat och begåvning (r_{23}). Följande korrelationer erhålles.

$$r_{12} = 0.50$$

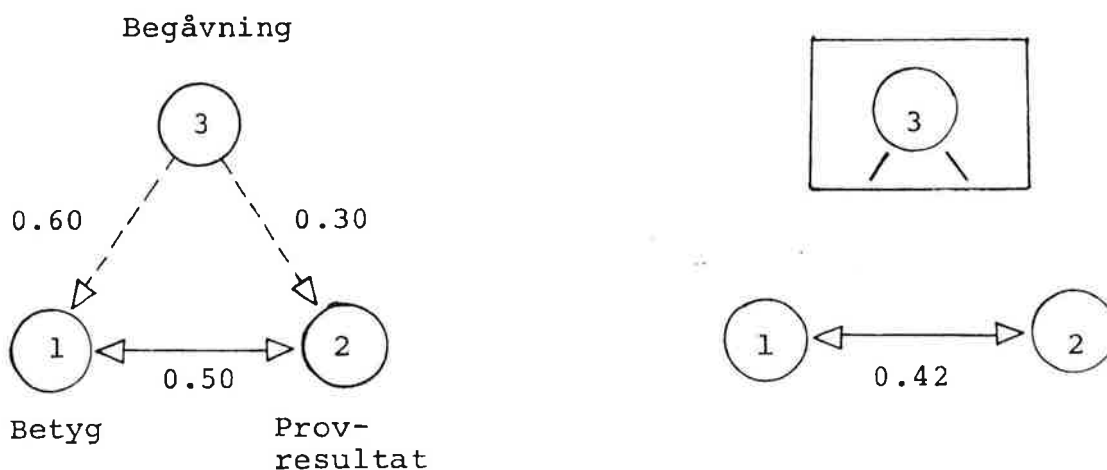
$$r_{13} = 0.60$$

$$r_{23} = 0.30$$

Den partiella korrelationen mellan slutbetyg och inträdesprov, när begåvning konstanthålles, blir:

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = \frac{0.50 - 0.60 \cdot 0.30}{\sqrt{(1-0.60^2)(1-0.30^2)}} = \frac{0.32}{0.76} = \underline{\underline{0.42}}$$

Resultatet belyses i följande figurer:



Resultatet innebär också att den gemensamma variansen i variabel 1 och 2 minskar från 25 % till 18 %.

Antag nu att man av någon orsak vill att begåvningen endast skall konstanthållas vad gäller resultaten i inträdesproven, men däremot får begåvningsvariabeln fritt påverka slutbetyget. I så fall beräknar man en semipartiell korrelation. Den skrivs $r_{1(2.3)}$, vilket innebär att man beräknar korrelationen mellan variabel 1 och 2, när inflytandet från variabel 3 enbart är eliminerat från variabel 2. Denna semipartiella korrelation beräknas med följande formel:

$$r_{1(2.3)} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{23}^2)}}$$

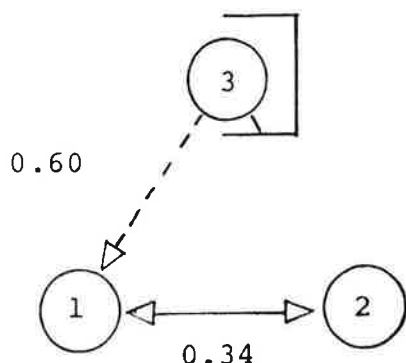
Denna formel liknar den vanliga formeln för beräkning av en partiell korrelation men termen $\sqrt{(1 - r_{13}^2)}$ saknas i nämnaren.

I vårt fall blir den semipartiella korrelationen

$$r_{1(2.3)} = \frac{0.50 - 0.60 \cdot 0.30}{\sqrt{(1 - 0.30^2)}} = \frac{0.32}{0.95} = 0.34$$

När begåvningen får "spela i full frihet" endast i betygsvariabeln, sjunker således korrelationen mellan betyg och inträdes-

prov till 0.34 och den gemensamma variansen blir endast 12 %.



Om man i stället hade velat att begåvningen enbart skulle ha konstanthållits i betygsvariabeln, hade den semipartiella korrelationen kunnat räknas fram med följande formel:

$$r_{2(1 \cdot 3)} = \frac{r_{12} - r_{23} \cdot r_{13}}{\sqrt{1 - r_{13}^2}}$$

Man kan även beräkna en semipartiell korrelation av andra ordningen, dvs. $r_{1(2 \cdot 34)}$. I detta fallet får man fram sambandet mellan variabel 1 och 2, när inflytandet från variabel 3 och 4 är eliminerat endast från variabel 2.

3. Förhållandet mellan semipartiell och multipel korrelation

Om vi har två okorrelerade prediktorer, erhålles den multipla korrelationen med kriteriet genom följande beräkning:

$$R^2 = r_{yx_1}^2 + r_{yx_2}^2$$

Om $r_{yx_1} = 0.50$, $r_{yx_2} = 0.50$ och $r_{x_1x_2} = 0$ blir

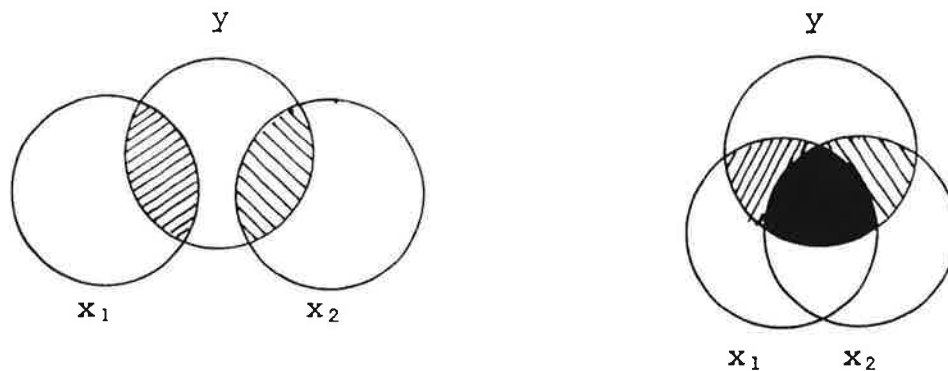
$$R^2 = 0.50^2 + 0.50^2 = 0.25 + 0.25 = 0.50$$

$$R = \sqrt{0.50} = 0.71$$

Den multipla korrelationen blir i detta fallet 0.71, vilket innebär att 50 % av variansen kan förklaras av variationen i de båda prediktorerna.

Om de båda prediktorerna däremot är korrelerade - vilket är det vanliga förhållandet - kan man ej göra samma enkla addition, eftersom prediktorerna ej längre bidrar med unik information. Det är delvis samma variation i kriteriet som förklaras av de båda prediktorerna.

I nedanstående figurer åskådliggörs prediktioner med okorrelerade respektive korrelerade x-variabler. Försök att förklara figurernas innebörd!



Även när prediktorerna är korrelerade kan man dock nå fram till den multipla korrelationen med hjälp av en addition. Formeln ser i detta fall ut så här:

$$R^2 = r_{YX_1}^2 + r_{Y(X_2 \cdot X_1)}^2$$

dvs. den förklarade variansen i kriterievariabeln = den kvadrerade korrelationen mellan y och x_1 + den kvadrerade semipartiella korrelationen mellan y och x_2 , där x_2 är opåverkad av x_1 . Den semipartiella korrelationen ger oss i detta fallet möjlighet att få information om, hur mycket en tillkommande prediktor ger i form av förklarad varians.

Vi återgår till exemplet med de tre variablerna betyg, begåvning och inträdesprov. Korrelationen mellan betyg och begåvning var 0.60 och den semipartiella korrelationen mellan betyg och provresultat 0.34. Vi vill nu använda oss av de båda variablerna begåvning och inträdesprov för att skatta slutbetyget, varför vi döper om betygsvariabeln till y, begåvningsvariabeln

till x_1 och provvariabeln till x_2 . Enligt den senast redovisade formeln blir den multipla korrelationen:

$$R^2 = 0.60^2 + 0.34^2 = 0.36 + 0.12 = 0.48$$

$$R = \sqrt{0.48} = 0.69$$

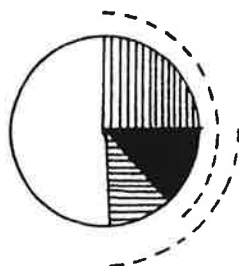
Om man beräknat den multipla korrelationen med tidigare givna formler, hade resultatet blivit detsamma.

$$\beta_1 = \frac{0.60 - 0.50 \cdot 0.30}{1 - 0.30^2} = 0.50$$

$$\beta_2 = \frac{0.50 - 0.60 \cdot 0.30}{1 - 0.30^2} = 0.35$$

$$R = \sqrt{0.50 \cdot 0.60 + 0.35 \cdot 0.50} = 0.69$$

Med hjälp av den nya formeln kan vi direkt se att 36 % av variansen i slutbetyget kan förklaras av begåvningsvariabeln. Använder vi oss dessutom av inträdesprovet ökar den förklarade variansen med 12 % upp till 48 %. Egentligen förklarar inträdesprovet 25 % av variansen i betyget ($r_{yx_2} = 0.50$) men 13 % av dessa 25 % har redan gottskrivits begåvningsvariabeln.



+++ 36 % av variansen i betyget förklaras utifrån begåvningskillnader

+++ 25 % - 13 % = 12 % är inträdesprovets tillskott

Om antalet prediktorer ökas till tre, får man upplysning om den förklarade variansen i kriterievariabeln genom en utvidgning av den tidigare redovisade formeln:

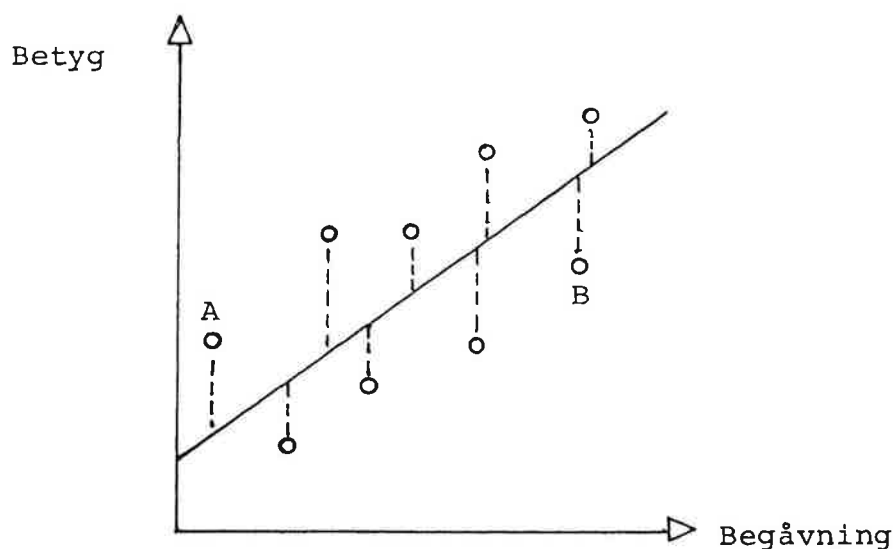
$$R^2 = r_{yx_1}^2 + r_{y(x_2 \cdot x_1)}^2 + r_{y(x_3 \cdot x_1 x_2)}^2$$

För varje tillkommande prediktor utökas formeln med en kvadrerad semipartiell korrelation. Genom formeln får man snabbt reda på varje nytillkommen prediktors unika bidrag, dvs. hur mycket av variansen i kriteriet som den nya och ingen av de tidigare prediktorerna kan förklara.

4. Alternativa beräkningssätt för partiell och semipartiell korrelation

För att ytterligare belysa sambandet mellan partiell, semipartiell och multipel korrelation tillåter vi oss att visa på en annan - om än ganska mödosam - väg att beräkna en partiell respektive en semipartiell korrelation. Vi utgår åter från exemplet med de tre variablerna betyg, begåvning och inträdesprov. I detta fallet ges inget sifferunderlag, utan vi väljer att endast ange det principiella tillvägagångssättet.

Vi börjar med att använda slutbetyget som kriterievariabel och begåvningen som prediktor. I ett koordinatsystem lägger vi in regressionslinjen, varefter de individuella avvikelserna till linjen uppmätes.

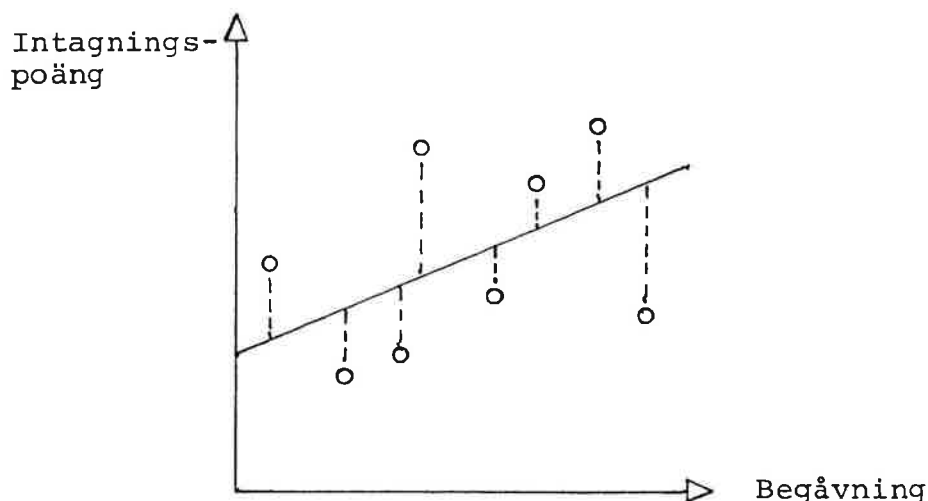


Avstånden till regressionslinjen upplyser oss om hur mycket högre eller hur mycket lägre betyg eleverna fått i förhållande till vad man skulle väntat sig utifrån deras begåvning. Med

hänsyn tagen till sambandet mellan begåvning och skolprestation har t.ex. individen A erhållit högre betyg än B, trots att B:s faktiska betyg är klart bättre. A:s relativa skolprestation är sålunda högre än B:s.

Avvikelserna från regressionslinjen, de s.k. residualerna, är nollkorrelerade med x-variabeln, dvs. det finns inte något samband mellan avvikelsernas tecken eller storlek och individernas poäng i begåvningsvariabeln.

Nu tar vi och byter kriterievariabel och ersätter betyget med intagningsprovet. I nedanstående koordinatsystem lägger vi in den aktuella regressionslinjen samt individernas avvikelser till denna. Liksom tidigare är dess residualer okorrelerade med x-variabeln.



Om vi nu beräknar korrelationen mellan residualerna i de båda figurerna, får vi veta vilket samband det finns mellan slutbetyg och intagningspoäng, när begåvningsinflytande har eliminerats i båda variablerna. (Båda avvikelsemåtten är ju okorrelerade med begåvningsvariabeln). Vad vi kommer fram till är alltså den partiella korrelationen mellan betyg och intagningspoäng med begåvningen konstanthållen och vi får exakt samma korrelationskoefficient (+0.42) som vi erhöll med tidigare redovisad formel. Detta betyder i sin tur att de individer som ligger över regressionslinjen i den ena figuren också tenderar att ligga över linjen i den andra. Likaså innebär det

att flertalet av de individer som hade lägre än väntade betyg utifrån begåvningsnivån också har lägre än predicerad intagningspoäng.

Om man i stället vill beräkna en semipartiell korrelation, t.ex. sambandet mellan betyg och inträdesprov, när inflytandet från begåvningsvariabeln är eliminerat endast från intagningspoängen, går man tillväga på följande sätt. Man korrelerar helt enkelt individernas avvikelser till regressionslinjen i den sist givna figuren med deras faktiskt erhållna betyg. På så vis har man eliminerat begåvningens inflytande från provresultatet men ej från betyget. Givetvis erhåller vi då samma korrelationskoefficient, som om vi använt den tidigare givna formeln.

Kapitel IV

KANONISK KORRELATION

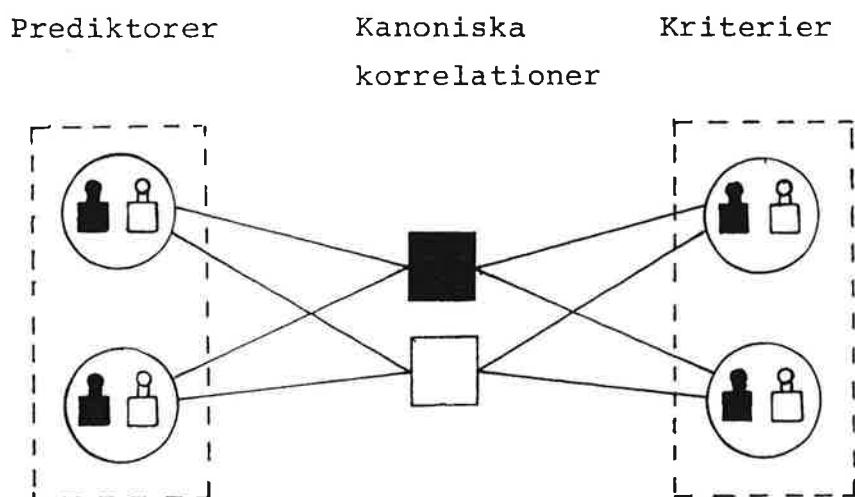
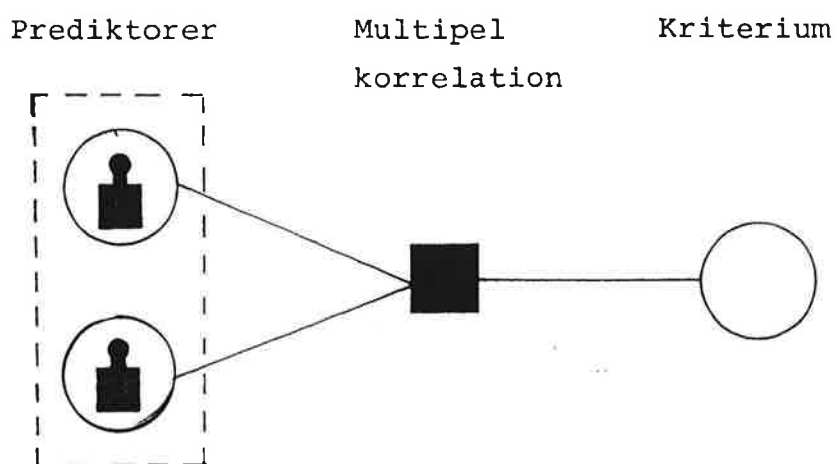
1. Introduktion

Om man har ett flertal prediktorer och ett kriterium, kan man beräkna en multipel korrelation. Har man även på kriteriesidan fler än en variabel, får man i stället göra en kanonisk korrelationsanalys. Vi skall i detta kapitel ge en kortfattad orientering om denna typ av sambandsmått.

De kanoniska korrelationerna kan betraktas som en typ av eller snarare en vidareutveckling av de multipla korrelationerna, eftersom de användes, när både prediktor och kriterium innehåller flera variabler. Vidare erhåller man inte en utan ett flertal korrelationskoefficienter, vars antal bl.a. är beroende av antalet enskilda variabler. I allmänhet minskar dock korrelationernas styrka mycket snabbt och jag har ej sett någon tillämpning av metoden som gett mer än tre signifikanta koefficienter.

Den första kanoniska korrelationskoefficienten erhålles, genom att de enskilda variablerna tilldelas vikter på ett sådant sätt att korrelationen maximeras mellan de sammanslagna variablerna. Man kan också uttrycka det så, att man har funnit en gemensam faktor eller komponent som svarar för huvudparten av samvariationen mellan de båda grupperna av variabler. Större delen av den återstående samvariationen kan förklaras av en andra gemensam faktor - nollkorrelerad med den första - som bestäms genom att variablerna tilldelas nya vikter, vilka uppfyller villkoret att korrelationen mellan de sammanslagna variablerna når ett maximum för denna faktor. Processen pågår till antalet kanoniska korrelationer når samma värde, som antalet variabler i den minsta av de båda variabeluppsättningarna.

Skillnaderna mellan multipel och kanonisk korrelation symboliseras i nedanstående figur.



2. Exemplifiering

Vi hämtar ett exempel på kanonisk korrelationsanalys från Svensson (1971). I denna undersökning ingår 8905 elever, vilka befann sig i årskurs sex 1961. Som prediktorer används här tre intelligenstest, ett verbalt test MOTSATSER, ett spatialt test PLÅTVIKNING samt ett induktivt test TALSERIER. Likaså har man tre kriterievariabler, betygen i LÄSNING, SKRIVNING och MATEMATIK:

Först beräknas den ovägda korrelationen mellan uppsättningarna av prediktorer och kriterier, dvs. det samband man får om man adderar råpoängen i de tre testen respektive i de tre

betygen för varje individ och beräknar en produktmomentkorrelation mellan dessa båda summor. Denna korrelation uppgår till 0.67, vilket innebär att 44.9 % av variansen i skolprestationsmättet kan förklaras utifrån den sammanslagna intelligenspoängen.

Härefter tillämpas tekniken med kanoniska korrelationer. Resultaten av beräkningarna finns i följande tabell, av vilken det framgår att den första kanoniska korrelationen uppgår till 0.73. Denna erhålles om poängen i testet Motsatser vägs med 0.95, poängen i testet Plåtvikning med 0.05 osv. Den andra kanoniska korrelationen uppgår till 0.33 och den tredje till 0.05. Trots att den tredje koefficienten är mycket låg, är även denna signifikant beroende på det stora antalet individer som ingår i analysen. (Signifikanstestningen görs med ett χ^2 -test).

Kanoniska korrelationer och viktkoefficienter

		Vikter för olika faktorer		
		1	2	3
Int. test	Motsatser	0.95	-0.63	-0.04
	Plåtvikning	0.05	0.25	-0.51
	Talserier	0.74	0.53	0.35
Betyg	Läsning	0.57	-0.50	-0.66
	Skrivning	0.38	-0.32	0.75
	Matematik	0.73	0.81	-0.08
Kanonisk korrelation		0.73	0.33	0.05

Med hjälp av den första faktorn i den kanoniska analysen kan man förklara 53.3 % ($0.73^2 \cdot 100$) av variansen i skolprestationsmättet utifrån skillnader i begåvningsmättet, vilket är närmare 10 % mer än vad som var möjligt med en enkel produktmomentkorrelation. Den andra faktorn förklarar ytterligare 10.8 % ($0.33^2 \cdot 100$) av den återstående oförklarade variansen. Sammanlagt förklarar dessa båda faktorer 53.3 % + 10.8 % (64.1 % av återstående 46.7 %). Den tredje faktorn bidrar föga och höjer endast den förklarade variansen med 0.1 %.

Om vi titar närmare på de tre faktorerna, finner vi att den första har relativt höga laddningar (vikter) i samtliga variabler med undantag för testet Plåtvikning. Den andra faktorn är bipolär med höga positiva laddningar i Talserier och Matematik samt negativa laddningar i Motsatser, Läsning och Skrivning. För en utförligare tolkning av faktorernas innebörd och dess betydelse för sambandet mellan begåvning och skolprestation hänvisas till originalrapporten.

Den kanoniska korrelationsanalysen kan också betraktas som en form av faktoranalys, en extern faktoranalys, där man försöker att finna de faktorer som kan förklara sambandet mellan två variabeluppsättningar. Denna externa faktoranalys ger en annan faktorstruktur än om man gör en faktoranalys inom vardera variabelgruppen. I det senare fallet gör man s.k. interna faktoranalyser som sprider ljus över sambanden inom grupperna.

Kapitel V

AVSLUTANDE SYNPUNKTER

Vi skall avsluta detta kortfattade kompendium med några kritiska synpunkter på de genomgångna metoderna samt ge en del litteraturreferenser.

Den multipla korrelationen ger oss ett mått på sambandet mellan en kriterievariabel och ett antal prediktionsvariabler. Genom att kvadrera den multipla korrelationskoefficienten får man veta, hur mycket av variansen i kriteriet som kan förklaras utifrån variansen i prediktorerna. Vidare kan man få kunskap om, hur stort bidrag som var och en av prediktorerna lämnar till den förklarade variansen och hur många av prediktorerna som ger ett signifikant tillskott. Utifrån den multipla korrelationen kan vi också genom multipel regression ställa en prognos för var och en av de enskilda individerna som ingår i analysen. Härvid kan vi använda oss av ett mycket stort antal prediktorer och vilken typ av prediktionsvariabler som helst. I en multipel regression är det t.ex. möjligt att blanda kontinuerliga och diskreta prediktionsvariabler (kön, civilstånd, socialgruppstillhörighet etc.). Den senare typen av variabler får emellertid transformeras till s.k. "dummy variables" (Se vidare Kerlinger & Pedhazur, 1973).

Beskrivningen ovan framställer den multipla regressionen i en mycket positiv dager och metoden har förvisso en mängd fördelar. I sanningens namn skall dock påpekas att den även är behäftad med en del nackdelar. En sådan är att den inbjuder till, att man mer eller mindre aningslöst kastar in en mängd prediktionsvariabler i förhoppning om, att åtminstone någon eller några - enskilt eller tillsammans - skall uppvisa signifikanta samband med kriteriet. Ett sådant förfarande är inte att rekommendera, därför att metoden är mycket känslig för regressionsvikternas reliabilitet. Härmed means att regressionskoefficienterna på grund av samplingsfel varierar från stickprov till stickprov, vilket innebär att de prediktorer eller grupper av prediktorer

som korrelerar högt med kriteriet i en analys, kanske inte alls gör det vid en förnyad utprovning. När man använder denna metod bör man därför klart motivera, vilka teoretiska eller praktiska utgångspunkter som styrt valet av den aktuella uppsättningen av prediktorer. Vidare bör man för att motverka alltför instabila regressionsvikter, arbeta med relativt stora stickprov - ett par hundra individer eller fler.

En annan nackdel med metoden är att den enskilde prediktorns bidrag till den förklarade variansen är starkt beroende av övriga prediktorer som ingår i analysen. En prediktor som ensam förklarar en stor del av variansen kanske blir helt betydelselös om man inför ytterligare en prediktionsvariabel. Detta beror på att den nya prediktorn är något högre korrelerad med kriteriet och dessutom har ett mycket högt samband med den första prediktorn. Problemet kan man dock komma tillrätta med genom olika tekniker, t.ex. genom att prioritera vissa prediktorers bidrag även om dessa är relativt små. En utförlig redogörelse för detta lämnas av Kerlinger & Pedhazur (1973, sid. 281) och Emanuelsson (1974, sid. 19).

Den partiella korrelationen används, när man vill studera sambandet mellan två variabler samtidigt som en eller flera andra variabler konstanthålles. Metoden är lätt att använda när man endast vill konstanthålla en variabel, men räknearbetet blir mycket mödosamt när de konstanthållna variablerna ökar i antal.

Den partiella korrelationen har således en kontrollerande funktion, medan den semipartiella korrelationen snarast har en förklarande funktion. Den semipartiella korrelationen intar nämligen en central plats inom den multipla regressionen och med hjälp av de kvadrerade semipartiella korrelationerna, kan vi direkt avläsa hur stort tillskott varje tillkommande prediktor ger i form av förklarad varians i kriteriet.

De kanoniska korrelationerna utnyttjas för att sprida ljus över sambandsstrukturen mellan två uppsättningar av variabler. Att beräkna de kanoniska korrelationerna manuellt är emellertid mycket tidsödande för att inte säga ogörligt när variabelan-

talet är stort. Det finns dock utmärkta dataprogram som snabbt levererar korrelationskoefficienter, regressionsvikter, signifikanstestningar och övriga önskade data. Vi vill emellertid uppmana till ett försiktigt handskande med metoden, när man använder sig av mycket små stickprov i kombination med stora variabeluppsättningar. I sådana fall kan man nämligen erhålla mycket höga korrelationer som endast baserar sig på slumpvariation (Gustafsson & Wernersson, 1975).

Vi skall nu övergå till att ge några litteraturhänvisningar. I de flesta större statistikböcker finns det något eller några kapitel som behandlar multipel och partiell korrelation. Som exempel kan nämnas:

Ferguson, G.A. (1971). *Statistical Analysis in Psychology and Education*. New York: MacGraw-Hill.

Harshbarger, T.R. (1971). *Introductory Statistics*. New York: MacMillan.

Walker, H.M. & Lew, J. (1953). *Statistical Inference*. New York: Holt, Rinehart & Winston.

Söker man en bok som utförligt och på ett relativt lättfattligt sätt redogör för en mångfald multivariata metoder, vill jag i första hand rekommendera:

Kerlinger, F.N. & Pedhazur, E.J. (1973). *Multiple Regression in Behavioral Research*. New York: Holt, Rinehart & Winston.

Ytterligare information om den multivariata problematiken lämnas av:

Cooley, W.W. & Lohnes, P.R. (1971). *Multivariate data analysis*. New York: Wiley.

Marriott, F.H.C. (1974). *The Interpretation of Multiple Observations*. London: Academic Press.

Mårdberg, B. (1969). *Dimensionsanalys för beteendevetare*. Stockholm: PA-rådet.

Som exempel på några undersökningar inom den pedagogiska forskningen, vilka utnyttjat multivariat metodik kan nämnas:

Emanuelsson, J. (1974). Social bakgrund, begåvning och utbildning som prediktorer för levnadsförhållanden i vuxen ålder. Rapport från pedagogiska institutionen vid lärarhögskolan i Stockholm.

Gustafsson, J-E. & Wernersson, I. (1975). Om multivariatanalys med få analysenheter och många variabler. Rapporter från pedagogiska institutionen, Göteborgs universitet.

Härnqvist, K. (1968). Relative changes in intelligence from 13 to 18. Scandinavian Journal of Psychology, 9, 50-82.

Svensson, A. (1971). Relative Achievement. Stockholm: Almqvist & Wiksell.

ÖVNINGSUPPGIFTER

1. Nedan framgår korrelationen mellan betyg efter avslutad utbildning vid Grönköpings badmästarskola (kriterievariabel) och ett test på grovmotoriken (prediktionsvariabel 1), korrelationen mellan det aktuella betyget och ett test på finmotoriken (prediktionsvariabel 2) samt korrelationen mellan de båda prediktorerna. Beräkna den multipla korrelationen.

	Y	x_1	x_2
Betyg (y)		.40	.35
Grovmotorik (x_1)			.20
Finmotorik (x_2)			

2. Försök att ge exempel på ett relativt högt samband mellan två variabler som helt eller nästan helt försvinner då man konstanthåller en tredje variabel.
3. I nedanstående tablå återfinns korrelationerna mellan längd, vikt och ålder för 103 pojkar i åldern 12 till 15 år:

	1	2	3
1. Längd		.70	.60
2. Vikt			.50
3. Ålder			

- a) Beräkna den partiella korrelationen mellan längd och vikt då åldern konstanthålles.
- b) Signifikanstesta denna partiella korrelation.
- c) Försök att tolka sambandet mellan längd och vikt.

4. I nedanstående tablå återfinns korrelationen mellan betyg i Svenska i årskurs 9 (kriterievariabel) och resultaten på ett ordförrådstest (prediktionsvariabel 1), korrelationen mellan det aktuella betyget och resultaten på ett läshastighetstest (prediktionsvariabel 2) samt korrelationen mellan de båda prediktorerna. I undersökningen ingår 203 elever.

	y	x_1	x_2
Betyg: Svenska (y)		.50	.40
Ordförråd (x_1)			.30
Läshastighet (x_2)			

- Vilka vikter skall de båda prediktorerna tilldelas, för att den vägda totalpoängen skall få maximalt samband med kriteriet?
 - Hur hög blir den multipla korrelationen?
 - Är denna korrelation signifikant?
5. För att kunna göra bästa möjliga selektion i en urvalssituation strävar man efter att få fram prediktorer som korrelerar högt med kriteriet, men som sinsemellan korrelerar lågt. Varför?
6. Två faktorer är av betydelse för hur många kilon en man orkar lyfta; hans egen vikt samt hur länge som han har tränat tyngdlyftning. I en studie fann man att korrelationen mellan vikt och lyftförmåga var 0.35, medan korrelationen mellan antalet träningstimmar och lyftförmåga uppgick till 0.70. Man undersökte också om det fanns något samband mellan kroppsvikt och träningstid och fann en korrelation på 0.15.
- Vad blir den multipla korrelationen om man väljer lyftförmåga som kriterievariabel samt vikt och träningstid som prediktorer?
 - Vad blir motsvarande samband om de båda prediktorerna tilldelas samma viktkoefficienter?

7. Ge en kortfattad redogörelse för följande begrepp:

- a) Residual
- b) Förklarad varians
- c) Extern faktoranalys
- d) Partiell korrelation av andra ordningen

8. I en undersökning mätte man konditionen bland 200 värnpliktiga. Dessa fick också uppge hur mycket kaffe de drack per dag. Man fann ett klart negativt samband mellan kondition och kaffekonsumtion. Korrelationskoefficienten uppgick till -0.60 . Undersökningen utsattes dock för kritik, därför att man inte hade tagit hänsyn till hur mycket de värnpliktiga rökte. I efterhand insamlades därför uppgifter om de värnpliktigas rökvanor, varefter man framräknade sambandet mellan kondition och tobakskonsumtion. Denna korrelation visade sig vara -0.70 . Sambandet mellan tobaks- och kaffekonsumtion befanns vara positivt, närmare bestämt $+0.80$. Med utgångspunkt från ovan relaterade data har Du att analysera kaffedrickandets menliga inverkan på konditionen.

9. Till ett träningsläger för tio-kampare väljer man ut 24 juniorer. Innan träningen börjar testar man ynglingarnas kondition samt deras muskelstyrka. Vid träningsperiodens slut anordnas en tävling i tio-kamp. Individernas resultat i denna tävling korreleras med den tidigare uppmätta konditionen respektive muskelstyrkan. De erhållna korrelationskoefficienterna redovisas i tabellen nedan.

	y	x_1	x_2
Tävlingsresultat (y)		.70	.60
Kondition (x_1)			.40
Muskelstyrka (x_2)			

- a) Hur hög blir den multipla korrelationen?
- b) Om man inför ytterligare en prediktor, ökar den multipla korrelationen med tre hundradelar. Ger den nya variabeln något signifikant tillskott till prediktionen?

10. I nedanstående tablå återfinns korrelationerna mellan poäng på ett skolmotivationstest, betyg i svenska samt resultat på ett intelligenstest för 95 elever i årskurs 9.

	1	2	3
1. Skolmotivation		.40	.30
2. Betyg i svenska			.60
3. Verbal begåvning			

- a) Beräkna den partiella korrelationen mellan skolmotivation och betyg i svenska då verbal begåvning konstanthålles.
- b) Beräkna den semipartiella korrelationen mellan motivation och betyg, då begåvningen endast konstanthålles i betygsvariabeln.
11. Redogör kortfattat för de väsentligaste skillnaderna mellan:
- a) multipel och partiell korrelation
- b) multipel och kanonisk korrelation
- c) partiell och semipartiell korrelation

12. I nedanstående tablå återfinns korrelationen mellan betyg i matematik i årskurs 8 (kriterievariabel) och resultaten på ett induktivt intelligenstest (prediktionsvariabel 1), korrelationen mellan det aktuella betyget och resultaten på ett spatialt intelligenstest (prediktionsvariabel 2) samt korrelationen mellan de båda prediktorerna.

	y	x ₁	x ₂
Betyg: Matematik (y)		.60	.40
Induktiv begåvning (x ₁)			.30
Spatial begåvning (x ₂)			

- a) Vilka vikter skall de båda prediktorerna tilldelas för att den vägda totalpoängen skall få maximalt samband med kriteriet?

- b) Hur hög blir denna maximala (multipla) korrelation?
- c) Vilka blir de väntade betygen för följande elever, om samtliga variabler är uttryckta i z-poäng?

	Resultat i intelligenstesten	
	Verbalt	Spatialt
Pelle	+1.0	+0.5
Greta	+0.5	-1.0
Fabian	-1.0	-1.5