

STOCKHOLMS UNIVERSITET  
Statistiska institutionen

**Analys av data från FIFA med hjälp av korrespondensanalys  
(Analysis of data from FIFA through correspondence analysis)**

Ximena Espinoza

15-högskolepoäng inom Statistik III, ht 2011  
Handledare: Göran Rundqvist

## Sammanfattning

Varje år utses världens bästa spelare av Federation Internationale de Football Association, FIFA, som grundar sig på röster från tränare, lagkaptener och journalister världen över. Denna uppsats använder sig av den här datan från åren 2008 till 2010, som omfattar diskreta variabler som spelarens ursprung och vilket land individerna som röstar på kommer ifrån. Korrespondensanalys kan ses som en sorts kanonisk korrelationsanalys. Den senare metoden analyserar relationen mellan kontinuerliga variabler medan korrespondensanalys analyserar förhållandet mellan diskreta variabler. Eftersom materialet består av data som inkluderar diskreta variabler har jag använt mig av korrespondensanalys som metod. Särskilt fokus riktas på korrespondensanalys men ett chi-två test används också för att fastställa om variablerna är oberoende av varandra. Huvudfrågan handlar om individernas preferens till en särskild fotbollsspelare och om denna påverkas av vilket land individerna kommer ifrån. Och om det finns något samband med spelarnas ursprung eller om kan man vara objektiv och rösta endast på prestation oavsett nationalitet.

<b>1. Inledning.....</b>	<b>4</b>
1.1 Syfte .....	4
1.2 Avgränsning .....	4
1.3 Disposition.....	4
<b>2. Datamaterial .....</b>	<b>5</b>
<b>3. Metod .....</b>	<b>6</b>
3.1 Chi-två test .....	6
3.2 Korrespondensanalys .....	6
3.3 Analys av data med hjälp av korrespondensanalys .....	7
3.4 Inertia .....	10
3.5 Koordinater.....	11
3.6 Punkternas inverkan till dimensionernas inertia .....	12
3.7 Dimensionernas inverkan till punkternas inertia .....	13
<b>4. Resultat.....</b>	<b>14</b>
4.1 Chi-två test .....	14
4.2 Korrespondensanalys .....	16
4.3 Grafisk resultat.....	20
<b>5. Diskussion.....</b>	<b>21</b>
<b>Litteraturförteckning.....</b>	<b>22</b>

# 1. Inledning

Genom loppet av sin 55-åriga historia har Ballon d`Or delats ut till några av de bästa spelarna som någonsin setts på spelplanen, från Stanley Matthews år 1956 till Lionel Messi under 2010. Ballon d`Or kan bokstavligen översättas till "den gyllene bollen" och är en internationell utmärkelse öppen för alla spelare, oavsett deras nationalitet eller liga där de spelar.

FIFA Ballong d`Or röstas fram av de nationella lagkaptener, tränare och en panel av fotbollsjournalister från hela världen. När trofén kom till var den ursprungligen för bästa europeiska fotbollsspelare som spelar i en europeisk liga, en regel som ändrades 1995, när icke-europeer som spelar för en av kontinentens klubbar blev också godtagbara. En ytterligare förändring skedde 2010 då utmärkelsen slogs samman med FIFA World player of the year.

Datamaterialet som ligger till grund till denna studie omfattar data från åren 2008 till 2010. Data för år 2010 skiljer sig från de andra två åren för att här har även journalister fått rösta. Under tidigare år har bara lagkaptener och tränare avgett röst. Datamaterialet är också uppdelat i ett system där man ger spelarna 5, 3 eller 1 poäng. Man röstar sammanlagt på tre spelare.

Metoden för studien är korrespondensanalys och målet är att undersöka om det finns ett samband mellan de olika variablerna: världsdel och spelarens ursprung/tillhörighet. Först används ett chi-två test för att fastställa om variablerna är oberoende av varandra och sedan analyseras data med hjälp av korrespondensanalys.

## 1.1 Syfte

Syftet med denna uppsats är att studera om det finns ett förhållande mellan de två olika variablerna.

Är sambandet mellan variablerna statistiskt signifikant?

Är individernas röster objektiva?

Påverkar ursprung vem man röstar på?

## 1.2 Avgränsning

Studien begränsas till åren 2008-2010 och till de manliga fotbollsspelarna. I denna uppsats tas bara upp om enkel korrespondensanalys.

## 1.3 Disposition

Kapitel två beskriver kortfattat om hur datamaterialet bearbetades. Vidare i kapitel tre ges teori om chi-två testet samt en introduktion till korrespondensanalys och hur man kan använda metoden för analys av data. Resultat kommer i kapitel 4 och diskussion i kapitel 5.

## 2. Datamaterial

Råmaterialet som ska analyseras kommer från FIFA:s hemsida [www.fifa.com](http://www.fifa.com) från åren 2008 till 2010. Urprugstabellerna är uppdelade i länder, namn på de som röstar och namn på tre fotbolls spelare som tilldelas 5, 3 respektive 1 poäng. Det finns två eller tre tabeller för varje år, en för tränarna, en annan för lagkapterna och en tredje för journalisterna. För att underlätta undersökningen delades länderna upp i världsdelar: Europa, Amerika, Afrika och Asien/Oceanien. Asien och Oceanien las ihop till en världsdel. Likaså fotbollsspelarnas nationaliteter delades upp i olika kontinenter: Europa, Amerika och Afrika. Världsdelarna och fotbollsspelarnas ursprung är de två olika variablerna.

Man fick slutligen fram en korstabell för till exempel hur många från Europa som röstar 3 poäng på en fotbollsspelare som kommer ursprungligen från Amerika o.s.v. Åren 2008 och 2010 sattes ihop till tre korstabeller: en för 5 poäng, den andra för tre poäng och en tredje för 1 poäng.

Contingency Table

	Eu	Am	Af	Sum
Europa	233	79	10	322
Amerika	119	73	11	203
Afrika	124	68	25	217
AsienOc	170	78	12	260
Sum	646	298	58	

Tabell 1 för 3-poäng rösterna för åren 2008-2010

## 3. Metod

### 3.1 Chi-två test

Chi-två testet används för data i nominalskala och testar oberoendet mellan till exempel två variabler i en korstabell. Testet kan användas för hypotestestning och mäter skillnad mellan observerad och förväntad frekvens.

Första steget i hypotestestning är att ange ett noll och ett alternativ hypotes.

Ho: Variablerna är oberoende av varandra

Ha: Variablerna är beroende av varandra

När vi vet hur många rader och kolumner vår tabell består av kan det kritiska värdet beräknas och för ett chi-två signifikanstest bestäms frihetsgraderna enligt följande formel:

$$df = (\text{antal rader}-1) (\text{antal kolumner}-1)$$

$$= (r-1) (c-1)$$

För till exempel ett test med 0.01 signifikansnivå och 3 frihetsgrader blir det kritiska värdet: 11.345. Beslutsregeln blir då att förkasta nollhypotesen om det beräknade värdet av chi-två är större än 11.345.

Vidare beräknas chi-två värdet enligt följande formel:

$$x^2 = \sum \left[ \frac{(f_o - f_e)^2}{f_e} \right]$$

där  $f_o$  är de observerade värden i varje cell och  $f_e$  är den förväntade värdet i cellerna. De förväntade frekvensvärdet i varje cell i tabellen beräknas på följande sätt:

$$f_e = \frac{(\text{radtotal})(\text{kolumntotal})}{\text{total}}$$

### 3.2 Korrespondensanalys

Korrespondensanalys är en teknik som studerar datamaterial som består av kategoriska variabler. Det finns två sorter korrespondensanalys: enkel korrespondensanalys (CA) och multipel korrespondensanalys (MCA). Syftet med korrespondensanalys är att visa strukturen som finns i ett komplext datamatrix utan att förlora information.

Korrespondensanalys gör det möjligt att presentera resultatet visuellt som punkter i ett flerdimensionellt plan för att på ett enkelt sätt tolka information.

Utgångspunkten för enkel korrespondensanalys är en tabell av två kategoriska variabler och

resultatet är en uppsättning koordinater för variabelernas kategorier, vilka på ett geometriskt sätt representerar hur dessa kategorier förhåller sig till varandra. Korrespondensanalys är med andra ord ett verktyg för att studera relationer mellan variabler. Kategorier som korrespondensanalys placerad nära varandra har liknande fördelningar och de punkter som placeras långt ifrån varandra har olika fördelningar.

Med hjälp av multipel korrespondensanalys kan man studera relationen mellan fler än två kategoriska variabler.

Korrespondensanalys har varit mycket populär i Frankrike sedan metoden utvecklades av den franska statistikern Jean-Pauls Benzecris under 1960-talet. Hans insats resulterade i publiceringen av ett stort arbete inom dataanalys. Analysen av ett system av relationen mellan rader och kolumner ledde till den metoden de kallade för korrespondensanalys.

Det arbetet som utförts av Benzecris och hans kollegor kan sägas representera en metodologisk filosofi byggd på induktiv resonemang.

I motsats till andra traditionella statistiska metoder som används för analys och tolkning av kategoriska variabler förutsätter inte den här metoden att det ska finnas en bestämd fördelning i materialet. Korrespondensanalys beskrivs ofta som en modell fri metod där materialet inte omfattas av många restriktioner, enda restriktionen är att variablerna inte får ha några negativa värden.

Korrespondensanalysens styrka ligger i att presentera en tabell det vill säga en matris på ett sätt så att relationen mellan rader och kolumner i tabellen blir möjliga att se direkt utan att behöva göra en mängd beräkningar i huvudet. En annan fördel med korrespondensanalys är att alla variabler kan användas, ingen särskild skal-nivå krävs utan allt behandlas som kategoridata. För tabeller med annat än frekvensdata måste man ibland göra transformationer till exempel för att inga negativa värden får ingå.

### **3.3 Analys av data med hjälp av korrespondensanalys**

Korrespondensanalys kan delas upp i tre steg:

1. Man börjar analysen med att beräkna profiler och radmassor för rader och kolumner som sedan delas upp i en rad respektive kolumntabell.
2. Sedan beräknar man avstånden mellan profilpunkterna.
3. Slutligen presenterar man resultatet genom att rita en graf.

För att beräkna profiler omvandlar man cellvärdena från den ursprungliga tabellen i proportioner, vilket görs för både rader och kolumner. Med andra ord divideras varje värde i den ursprungliga radtabellen med respektive rad-kolumnsumma. Detta innebär att två nya tabeller skapas: radprofil och kolumnprofil.

### Contingency Table

	Eu	Am	Af	Sum
Europa	179	134	5	318
Amerika	86	102	10	198
Afrika	134	73	13	220
AsienOc	144	113	2	259
Sum	543	422	30	995

Tabell 2. Korstabell över världsdela/ursprung för 5-poäng.

Man kan här visa som exempel radprofilen för Europa: om man dividerar värdena i europaraden 179, 134 och 5 med radsumman 318 får man följande:  $179/318=0.56$ ,  $134/318=0.42$  och  $5/318=0.016$ .

Radprofilerna motsvarar den relativa frekvensfördelningen för spelarens ursprung för varje världsdela.

Kolumnprofilerna räknas på liknande sätt där man dividerar värdena i första kolumnen 179, 86, 134 och 144 med kolumnsumman 543 och vi får:  $179/543=0.33$ ,  $86/543=0.16$ ,  $134/543=0.25$  och  $144/543=0.27$

### Row Profiles

	Eu	Am	Af
Europa	0.56	0.42	0.016
Amerika	0.43	0.52	0.05
Afrika	0.61	0.33	0.059
AsienOc	0.56	0.43	0.008

Tabell 3: Radprofiltabell

### Column Profiles

	Eu	Am	Af
Europa	0.33	0.32	0.17
Amerika	0.16	0.24	0.33
Afrika	0.25	0.17	0.43
AsienOc	0.27	0.27	0.07

Tabell 4: Kolumnprofiltabell

Sedan beräknar man de genomsnittliga radprofilerna genom att dividera total antal observationer i kolumnerna: 543, 422 och 30 med totalsumman 995. Vilket ger följande:  $543/995=0.54$ ,  $422/995=0.43$  och  $30/995=0.03$ .

De genomsnittliga radprofilerna motsvarar profilen av den marginella fördelningen av kolumnvariabeln (ursprung). Här kan man se att 54% av de som röstar, röstar på spelare som tillhör Europa.



Genomsnittliga radprofilerna är den genomsnittliga vikten av radprofilerna och den här punkten kallas ofta för centroiden och är grafens mittpunkt origo. Om en profil är väldigt olik genomsnittsprofilen, då kommer punkten att ligga långt ifrån origo, medan profiler som är nära genomsnittet kommer att representeras med punkter nära centroiden. Om alla kategorier har liknande profiler kommer alla punkter att hamna i centroiden. (Clausen 1998)

Radmassorna består av den relativa frekvensfördelningen för radsumman. (marginal fördelning). Om man fortsätter med samma exempel då får man  $318/995=0.32$ ,  $198/995=0.20$ ,  $220/995=0.22$ ,  $259/995=0.26$ .

	Quality	Mass	Inertia
Europa	1.0000	0.3196	0.0789
America	1.0000	0.1990	0.3688
Africa	1.0000	0.2211	0.4040
AsienOc	1.0000	0.2603	0.1484

Tabell 5. Summary Statistics for the Row Points

Eftersom varje radprofil är oberoende av det totala antalet i den berörda kategorin används radmassorna för att lämna information om detta nummer. De kan sägas vara ett mått på betydelsen av en viss profil på analysen.

Varje radprofil ska betraktas som en matematisk vektor och vektorn ska representeras som en punkt i ett flerdimensionellt plan där varje profil i utgör en koordinat i rummet. För att få en grafisk presentation och för att rita ut kontinenternas position av avståndet mellan dem måste man behandla radvärdena som koordinater i ett rum. Ju närmare varandra världsdelarna hamnar, desto mer liknar dem varandra. En fördel med korrespondensanalys är att man kan rita in punkterna för rader och kolumner i samma graf.

I korrespondensanalys använder man begreppet chi-två avståndet för att beräkna avstånden mellan de olika punkterna. För att beräkna avstånden kan man använda följande formel:

$$d(i, i') = \sqrt{\sum_j \frac{(a_{ij} - a_{i'j})^2}{a_{.j}}}$$

Där  $d(i, i')$  är chi-två avståndet mellan punkterna  $i$  och  $i'$ ,  $a_{ij}$  är värdena i profilerna och  $a_{.j}$  är värden i den genomsnittliga profilerna.

Genom att använda värdena i radprofilerna i tabell 2 och de uträknade genomsnittliga radprofilerna kan vi till exempel beräkna chi-två avståndet mellan Europa och Amerika.

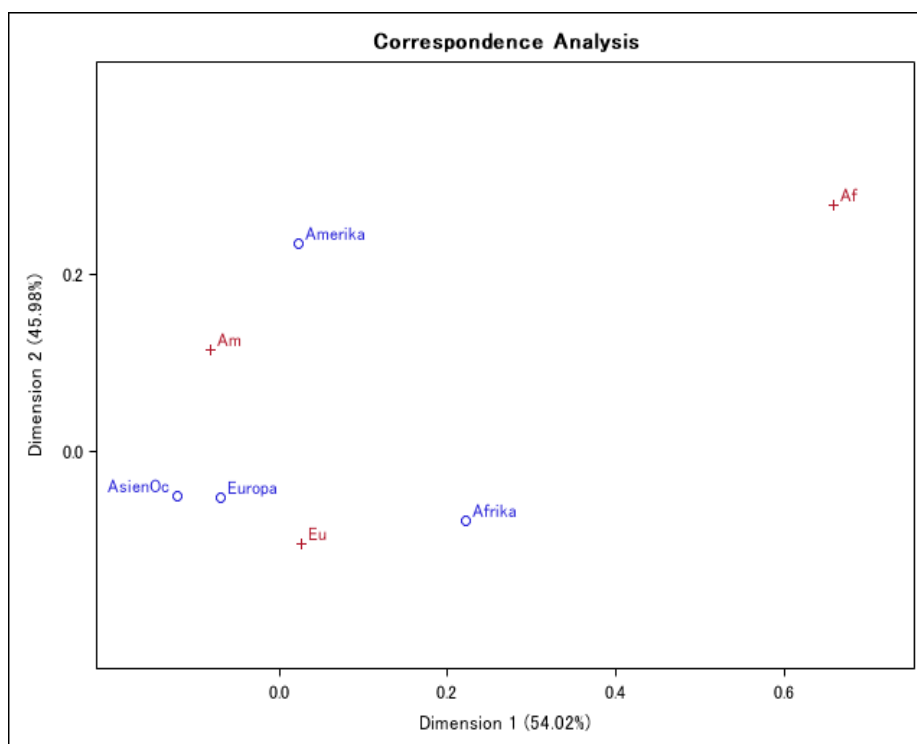
$$d(1,2) = \sqrt{\frac{(0.56 - 0.43)^2}{0.54} + \frac{(0.42 - 0.515)^2}{0.42} + \frac{(0.0157 - 0.05)^2}{0.033}} = 0.301$$

Täljaren innehåller, för varje spelare ursprung, de kvadrerade skillnaderna mellan profilerna Europa och Amerika, dividerat med de genomsnittliga profilerna för respektive ursprung av spelare.

Det är viktigt att påpeka att avstånden kan bara beräknas mellan kategorier av samma variabel, inte mellan kategorier av olika variabler. I det här fallet kan man alltså beräkna avstånden mellan världsdelarna till exempel Europa och Amerika, men inte mellan världsdelar och spelarens ursprung.

När liknande beräkningar har gjorts för kolumner är det möjligt att rita ut punkterna i ett flerdimensionellt plan.

Nästa steg att rotera axeln så att den ska ligga så nära som möjligt till punkterna. Avståndet mellan världsdelarna och samtliga spelare ska stämma så bra som möjligt med spelarnas kolumn i radprofilerna.



Figur 1. Tvådimensionell karta över olika världsdelar och spelarens ursprung, 5-poängstabellen

### 3.4 Inertia

I korrespondensanalys är begreppet inertia synonym till varians och den hänger ihop med chi-två avstånden. Ett mått på hur profilpunkterna är spridda runt centroiden kallas för total inertia och beräknas med följande formel:

$$A^2 = \sum_i r_i d_i^2$$

där  $d_i$  är punkten  $i$ 's chi-två avstånd från centroiden och  $r_i$  är punkten  $i$ 's massa (vikten). Total inertia är direkt relaterade till Pearsons chi-två statistiska enligt följande:

$$x^2 = A^2 N$$

där  $N$  är antal observationer i tabellen.

Man kan också observera från ovanstående ekvation att den totala inertia ( $A^2$ ) är lika med Pearsons kontingents koefficient eller phi i kvadrat.

$$\phi^2 = \frac{x^2}{N} = A^2$$

Korrespondensanalys kan också beskrivas som en teknik för decomposition av phi i kvadrat för en frekvenstabell.

Den totala inertia är uppdelad av ett set av egenvärden där varje axels bidrag till variansen kallas axelns eller dimensionens egenvärde.

Egenvärdena visar hur stor del av den totala inertia varje dimension förklarar. Första dimensionen förklarar mest och sedan den andra o.s.v.

Singular Value	Principal Inertia	Chi-Square	Chi-Percent	Cumulative Percent	11	22	33	44	55
0.12778	0.01633	16.2464	54.02	54.02	*****				
0.11789	0.01390	13.8276	45.98	100.00	*****				
Total	0.03023	30.0740	100.00						

Degrees of Freedom = 6

### Tabell 6. Inertia and Chi-square decomposition

I den här datan behövs det två dimensioner för att förklara den totala inertian. Tabellen 4 visar egenvärden för vår datamatrix. Man kan se att första dimensionen förklarar 54.02 av inertian och dimension 2 förklarar resterande inertia 45.98%.

Singular values används här istället för egenvärden. Singular values är kvadratroten ur egenvärdena.

När lösningen förklarar 100% av inertian då är avstånden mellan punkterna inom varje set av punkter lika med chi-två avståden. Om man analyserar större tabeller är det ofta så att bara en mindre del av inertian är förklarad och i detta fall kan avstånden bara bli approximativt till chi-två avstånden.

### 3.5 Koordinater

Koordinaternas värde visar punkternas position i förhållandet till dimensionerna. Man kan tolka resultatet på basis av de relativa punktpositionerna och deras fördelning längs dimensionerna. (Clausen 1998)

#### Row Coordinates

	Dim1	Dim2
Europa	-0.0690	-0.0520
Amerika	0.0229	0.2356
Afrika	0.2218	-0.0776
AsienOc	-0.1212	-0.0503

#### Tabell 7. Radkoordinater

Vi har tidigare använt d2 för att beräkna den totala inertia och vi kan på liknande sätt använda koordinaternas punkter för att hitta egenvärdena av varje dimension. Vi kan använda följande formeln:

$$\lambda_k^2 = \sum_i r_i f_{ik}^2$$

Där  $f_{ik}^2$  är koordinaternas punkt i kvadrat i dimensionen k och  $r_i$  är punkten i's massa. För första dimensionen ges:

$$\lambda^2 = 0.32(-0.07^2) + 0.20(0.02^2) + 0.22(0.22^2) + 0.26(-0.12^2) = 0.016$$

### 3.6 Punkternas inverka till dimensionernas inertia

För att få en bättre och korrekt tolkning av resultaten av korrespondensanalys kan man använda "Punkternas inverka till dimensionernas inertia" och "Dimensionernas inverka till punkternas inertia"

Punkternas inverka till dimensionernas inertia kan tolkas som proportionen av inertia från en särskild dimension som förklaras av punkten och visar i vilken utsträckning punkten har bidragit till att bestämma riktningen av den berörda dimensionen. (Clausen 1998) Detta är en hjälp för tolkningen av dimensioner och punkter med stora fördelningar och är mycket viktiga för den berörda dimensionen. Inom varje set av poäng är summan av dessa bidrag för varje dimension lika med ett.

Punkternas inverka till dimensionerna definieras med följande formel:

$$ca_{ik} = \frac{r_i f_{ik}^2}{\lambda_k^2}$$

där  $r_i$  är massan av profilpunkten  $i$ ,  $f_{ik}$  är punktens koordinat i dimensionen  $k$  och  $\lambda_k^2$  är dimensionens  $k$  egenvärde.

### 3.7 Dimensionernas inverkan till punkternas inertia

Nästa steg i tolkningen av resultaten är att bestämma hur väl varje punkt är förklarad av varje dimension. Detta visar sig i vilken utsträckning varje dimension bidrar till punkternas inertia vilket ger information om hur mycket punkternas inertia är förklarad av varje dimension.

(Clausen 1998)

Detta kan beräknas med följande formel:

$$cr_i = \frac{f_i^2}{d_i^2} = \cos^2 \omega$$

Dimensionernas inverkan till punkterna är oberoende av punkternas massor och visar hur väl punkten är förklarad av dimensionen i fråga. Den här inverkan kallas ofta för korrelation i kvadrat eftersom det är samma som  $\cos^2$  för vinkeln mellan linjen från centroiden till punkten och huvudaxlarna. Om  $cr$  för en dimension är hög då är vinkeln mellan punktens vektor och axel liten.

Punkten är därför belägen i axelns riktning, vilket innebär en stor korrelation med dimensionen. Summan av  $cr$  värden för varje profilpunkt över alla dimensioner är lika med ett. När man analyserar stora tabeller, då är det bara ett antal dimensioner som är intressanta. I sådana fall, visar summan av korrelationen i kvadrat godheten i testet (goodness of fit) i varje punkts representation i lösningen och är ofta kallad kvaliteten av varje punkts beskrivning. Denna statistiska motsvarar communalities i principal component analysis.

Det är viktigt att förstå skillnaden mellan de två olika bidragen. Punkternas bidrag till dimensionernas inertia fungerar först och främst som vägledning för tolkningen av dimensionerna och korrelationen i kvadrat visar hur väl en punkt förklaras av dimensionen. Oftast om punkternas inverkan till dimensionerna är höga kommer korrelationen i kvadrat också att vara hög, men inte tvärtom. Eftersom båda de här statistiska är alltid positiva är det viktigt att granska koordinaterna för att se i vilken riktning i dimensionen punkten finns.

## 4. Resultat

### 4.1 Chi-två test

För att ta reda på om variablerna i datamaterialet är oberoende av varandra används ett chi-två test. Nollhypotesen i denna test är att det inte finns något samband mellan variablerna. Man gör ett test för alla tre korstabeller.

Contingency Table

	Eu	Am	Af	Sum
Europa	179 (173.54)	134 (134.9)	5 (9.59)	318
Amerika	86 (108.05)	102 (83.98)	10 (5.97)	198
Afrika	134 (120)	73 (93.3)	13 (6.63)	220
AsienOc	144 (141)	113 (109.84)	2 (7.2)	260
Sum	543	422	30	

Tabell 8 Observerade värden och förväntad värde för 5-poängstabellen.

Förväntade värden beräknas och till exempel för Europa/Eu blir resultatet följande:

$$f_e = \frac{(\text{radtotal})(\text{kolumntotal})}{\text{total}} = 318 \times 543 / 995 = 173.54$$

Nu kan man räkna chi-två värdet:

$$x^2 = \sum \left[ \frac{(f_o - f_e)^2}{f_e} \right] = \frac{(179 - 173.54)^2}{173.54} + \dots + \frac{(2 - 7.8)^2}{7.8} = 30$$

Eftersom vi vet att tabellen har 4 rader och tre kolumner då blir frihetsgraderna:

$$df = (r-1)(c-1) = (4-1)(3-1) = 6$$

För ett test med 0,05 signifikansnivå och 6 frihetsgrader blir det kritiska värdet: 12.6 Och eftersom  $30 > 12.6$  då förkastas nollhypotesen. Med andra ord variablerna i 5-poängstabellen beroende av varandra och det finns ett samband mellan dem.

Man kan beräkna på samma sätt för 3 och 1 poängstabellerna:

**Contingency Table**

	<b>Eu</b>	<b>Am</b>	<b>Af</b>	<b>Sum</b>
<b>Europa</b>	233 (207. 6)	79 (95. 76)	10 (18. 63)	322
<b>Amerika</b>	119 (130. 88)	73 (60. 37)	11 (11. 75)	203
<b>Afrika</b>	124 (139. 90)	68 (64. 54)	25 (12. 56)	217
<b>AsienOc</b>	170 (167. 62)	78 (77. 33)	12 (15. 05)	260
<b>Sum</b>	646	298	58	

Tabell 9. Observerade värden och förväntad värde för 3-poängstabellen.

**Contingency Table**

	<b>Eu</b>	<b>Am</b>	<b>Af</b>	<b>Sum</b>
<b>Europa</b>	221 (200. 65)	85 ( 84. 5)	13 (33. 81)	319
<b>Amerika</b>	131 (127. 06)	55 (53. 53)	16 (21. 4)	202
<b>Afrika</b>	112 (137. 75)	55 (58. 04)	52 (23. 21)	219
<b>AsienOc</b>	165 (163. 54)	70 (68. 9)	25 (27. 56)	260
<b>Sum</b>	629	265	106	

Tabell 10. Observerade värden och förväntad värde för 1-poängstabellen

Chi-två-värdet för 3 och 1 poängstabellen blir följande:

$$x^2 = \sum \left[ \frac{(f_o - f_e)^2}{f_e} \right] = 28.78$$

$$\chi^2 = \sum \left[ \frac{(f_o - f_e)^2}{f_e} \right] = 42.45$$

28.78 > 12.6 och 42.45 > 12.6 Både för tre- och en poängstabellen förkastas nollhypotesen, variablerna är beroende av varandra och det finns ett samband mellan dem.

## 4.2 Korrespondensanalys

Med hjälp av korrespondensanalys vill man undersöka om det finns något samband mellan de två olika variablerna i materialet. Tabell 11 visar Inertia som motsvarar variansen och egenvärdena. Den totala inertia är summan av egenvärdena.

### Inertia and Chi-Square Decomposition

Singular Value	Principal Inertia	Chi-Square	Percent	Cumulative Percent	16	32	48	64	80
0.15384	0.02367	23.7138	82.39	82.39	-----+-----+-----+-----+-----				
0.07113	0.00506	5.0694	17.61	100.00	*****				
Total	0.02873	28.7832	100.00						

Degrees of Freedom = 6

Tabell 11. Inertia och chi-square decomposition för 3-poängtabellen

Egenvärdena visar hur mycket av den totala inertia varje faktor förklarar. Singular values är kvadratroten av egenvärdena. Här kan man läsa att dimension 1 förklarar 82.39 procent av variansen och dimension 2 förklarar 17.61 procent av variansen.

### Inertia and Chi-Square Decomposition

Singular Value	Principal Inertia	Chi-Square	Percent	Cumulative Percent	20	40	60	80	100
0.23931	0.05727	57.2674	99.86	99.86	-----+-----+-----+-----+-----				
0.00882	0.00008	0.0777	0.14	100.00	*****				
Total	0.05735	57.3452	100.00						

Degrees of Freedom = 6

Tabell 12 Inertia och chi-square decomposition för 1-poängtabellen

Tabell 12 visar att dimension 1 är väldigt dominant och förklarar 99.86 procent av variansen medan dimension 2 förklarar 0.14 procent av variansen.



Tabell 13 visar bland annat radpunkternas massor. Radmassorna är ett mått på punktens relevans i analysen och massor med höga värden visar vilka kontinenter inkluderar flest röster. Objekt med tung massa påverkar i högre grad centroidens placering än objekt med mindre massa.

**Summary Statistics for the Row Points**

	<b>Quality</b>	<b>Mass</b>	<b>Inertia</b>
<b>Europa</b>	1.0000	0.3214	0.3491
<b>Amerika</b>	1.0000	0.2026	0.1309
<b>Afrika</b>	1.0000	0.2166	0.4972
<b>AsienOc</b>	1.0000	0.2595	0.0228

Tabell 13. Summeringsstatistik över radpunkterna för 3-poängs korstabell.

I 3-poängstabellen har objektet Europa en tung massa = 0.3214.

**Summary Statistics for the Row Points**

	<b>Quality</b>	<b>Mass</b>	<b>Inertia</b>
<b>Europa</b>	1.0000	0.3190	0.2594
<b>Amerika</b>	1.0000	0.2020	0.0267
<b>Afrika</b>	1.0000	0.2190	0.7092
<b>AsienOc</b>	1.0000	0.2600	0.0047

Tabell 14. Summeringsstatistik över radpunkterna för 1-poängs korstabell.

Europa har tyngst massa med 0.3190.

Om vi undersöker radpunkternas inverkan på dimension 1 och dimension 2 som visas i tabell 15, kan man se att Europa, Afrika och AsienOc bidrar mest till den första dimensionen (9.3%, 66% och 23%), medan Amerika bidrar mest till dimension 2 (79%).

**Partial Contributions to Inertia for the Row Points**

	<b>Dim1</b>	<b>Dim2</b>
<b>Europa</b>	0.0931	0.0622
<b>Amerika</b>	0.0064	0.7946
<b>Afrika</b>	0.6663	0.0958
<b>AsienOc</b>	0.2343	0.0474

Tabell 15. Partiell inverkan av inertian på radpunkterna för 5-poängs korstabell.

#### Partial Contributions to Inertia for the Row Points

	Dim1	Dim2
Europa	0.3918	0.1492
Amerika	0.0324	0.5916
Afrika	0.5553	0.2254
AsienOc	0.0205	0.0338

Tabell 16. Partiell inverkan av inertian på radpunkterna för 3-poängs korstabell.

Europa och Afrika bidrar mest till dimension 1 medan Amerika och AsienOc bidrar mest till dimension 2.

#### Partial Contributions to Inertia for the Row Points

	Dim1	Dim2
Europa	0.2592	0.4195
Amerika	0.0263	0.3230
Afrika	0.7101	0.0610
AsienOc	0.0044	0.1964

Tabell 16. Partiell inverkan av inertian på radpunkterna för 1-poängs korstabell.

Afrika bidrar mest till dimension 1 medan Europa, Amerika och AsienOc bidrar mest till dimension 2.

#### Indices of the Coordinates That Contribute Most to Inertia for the Row Points

	Dim1	Dim2	Best
Europa	0	0	1
Amerika	0	2	2
Afrika	1	1	1
AsienOc	1	0	1

Tabell 17. Indikationer på koordinater som mest bidrar till inertian för radpunkterna för 5-poängs korstabell.

På samma sätt kan man se i ovanstående tabell att Europa, Afrika och AsienOc bidrar bäst till dimension 1 vilket stämmer bra med föregående uträkningar.

#### Squared Cosines for the Row Points

	Dim1	Dim2
Europa	0.6374	0.3626
Amerika	0.0093	0.9907
Afrika	0.8910	0.1090
AsienOc	0.8531	0.1469

Tabell 18. Squared Cosines for the row points, 5-points table.

Tabell 18 visar dimension 1 och dimension 2 inverkan på inertia på radpunkterna. Dimension 1 påverkar placeringen av objektet Europa med 64% och dimension 2 med 36%.

**Squared Cosines for the Row Points**

	<b>Dim1</b>	<b>Dim2</b>
<b>Europa</b>	0. 9247	0. 0753
<b>Amerika</b>	0. 2037	0. 7963
<b>Afrika</b>	0. 9202	0. 0798
<b>Asien0c</b>	0. 7392	0. 2608

Tabell 19. Squared Cosines for the row points, 3-points table.

Dimension 1 påverkar placeringen av objektet Europa med 92% och dimension 2 med 7%.

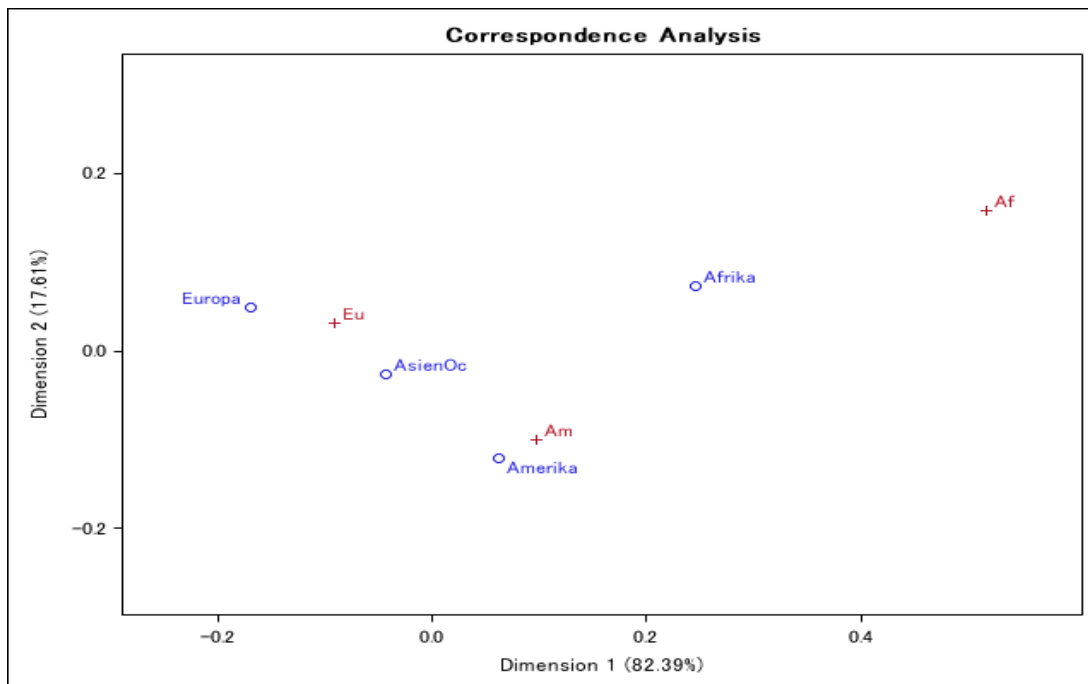
**Squared Cosines for the Row Points**

	<b>Dim1</b>	<b>Dim2</b>
<b>Europa</b>	0. 9978	0. 0022
<b>Amerika</b>	0. 9836	0. 0164
<b>Afrika</b>	0. 9999	0. 0001
<b>Asien0c</b>	0. 9431	0. 0569

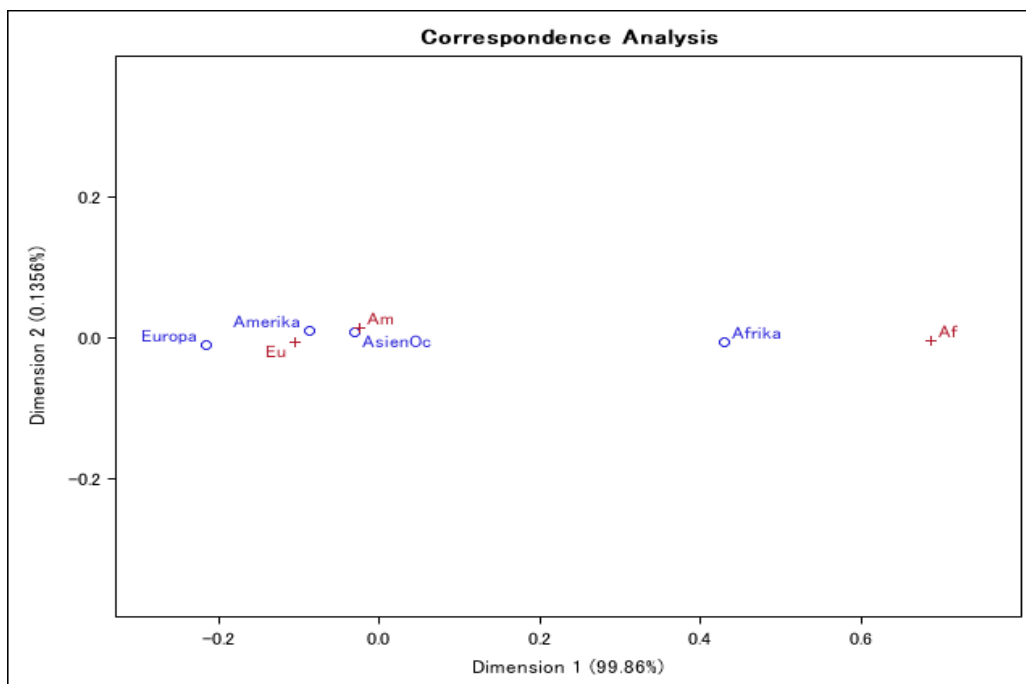
Tabell 20. Squared Cosines for the row points, 1-point table.

Dimension 1 påverkar placeringen av objektet Europa med 99.78 % och dimension 2 med 0.002 %.

### 4.3 Grafisk resultat



Figur 2. Tvådimensionell karta över olika världsdelar och spelarens ursprung, 3-poängstabellen



Figur 3. Tvådimensionell karta över olika världsdelar och spelarens ursprung, en poängstabell.

## 5. Diskussion

Benzecri var kritisk mot chi-två testet och använde istället tolkningsbarheten som ett kriterium. Andra rekommenderar att man gör ett chi-två test på tabellen för att fastställa om det finns något samband att analysera. I den här undersökningen gjordes det ett chi-två test på samtliga korstabeller och alla tester visade sig vara signifikanta, med andra ord så finns det en relation mellan variablerna.

Vidare användes korrespondensanalys för att på ett lätt och visuellt sätt visa i ett flerdimensionellt plan om det fanns ett samband mellan variablerna.

Tolkningen i konfigurationen av punkterna grundar sig på chi-två avstånden mellan punkterna och dessa avstånd är definierade separat för varje set av punkter. Detta innebär att om två rader punkter ligger nära varandra så ligger profilerna av de här två punkterna nära varandra. Ju mer olika profilerna är varandra desto längre bort från varandra kommer punkterna att hamna.

Avståndet mellan två punkter av olika slag till exempel mellan världsdel och spelarens ursprung har ingen mening i sig. Men avstånden mellan en världsdel och två ursprungskontinenter, dvs skillnaden mellan avståndet mellan en världsdel och en ursprungskontinent och avståndet mellan samma världsdel och en annan ursprungskontinent är meningsfull.

Om man tittar på figur 1 då kan man fråga sig:

Vad betyder det då att avståndet mellan Europa och Eu är mindre än avståndet mellan Europa och Am och mindre än avståndet mellan Europa och Af? Det betyder att i jämförelse med Am och Af är andelen på de som röstar på Eu från Europa, jämfört med alla som röstar på Eu är större. Det behöver inte betyda att det är vanligast att man röstar på en spelare från Europa om man kommer från Europa. (mätt i absoluta tal).

I det här fallet för 5-poängstabellen till exempel så var det flest europeer som röstade på europeiska spelare, man skulle alltså kunna dra den slutsatsen. Det var också flest individer från kontinenten Amerika som röstade på spelare från samma kontinent. Man skulle kunna säga att ursprung påverkar vem man röstar på men det är svårt att dra generella slutsatser.

## Litteraturförteckning

Böcker:

Clausen S. E. (1998). Applied correspondence analys: an introduction, Thousand Oaks, sage publications.

Susan C. Weller A. Kimball Romney (1990). Metric Scaling, correspondence analys, Newbury Park, sage publications.

Elektroniska källor:

[www.skeptron.uu.se/pers/carinac/praxeology/Rapport\\_4.pdf](http://www.skeptron.uu.se/pers/carinac/praxeology/Rapport_4.pdf)

[http://sociologi.cjb.net/vetenskapliga\\_metoder/korrespondens\\_analys.pdf](http://sociologi.cjb.net/vetenskapliga_metoder/korrespondens_analys.pdf)

[www.fifa.com/ballondor/news/newsid=1528255](http://www.fifa.com/ballondor/news/newsid=1528255)

[www.fifa.com](http://www.fifa.com)