Korrelation.*)

Af

Edv. Ph. Mackeprang.

Statistikeren opstiller sine Tal i Tabeller af forskellig Form, dels i Tabeller med enkelt, dels i Tabeller med dobbelt Indgang. Et Eksempel paa en Tabel med dobbelt Indgang findes i Befolkningsstatistiken, naar Brudgommens og Brudens Aldre stilles i Forhold til hinanden, jfr. omstaaende Tabel I, medens f. Eks. en Tabel med enkelt Indgang kun angiver Brudgommenes Antal i forskellige Aldersklasser, jfr. Tabel II. Man kan sammenligne de to Tabelformer henholdsvis med en grafisk Fremstilling i Rummet og i Planen; i sidstnævnte kan der kun afsættes ud ad 2 Akser, man kan kun fremstille x og y's indbyrdes Afhængighed, i først-

*) Den væsentligste Litteratur angaaende dette Emne: Bowley: Elements of Statistics, London 1901, p. 316—328. Davenport: Statistical methods with special reference to biological variation, New York 1899. Duncker: Die Methode der Variationsstatistik, Leipzig 1899. Norton: Statistical studies in the New York money-market, New York 1902. Pearson: Mathematical contributions to the theory of evolution i Philos. Transact. Roy. Soc. London, 1894, 1895 og 1896, samt talrige andre Afhandlinger i samme. Yule: On the theory of correlation i Journal of the Roy. Stat. Society, 1897.

Korrelation.*)

Af

Edv. Ph. Mackeprang.

Statistikeren opstiller sine Tal i Tabeller af forskellig Form, dels i Tabeller med enkelt, dels i Tabeller med dobbelt Indgang. Et Eksempel paa en Tabel med dobbelt Indgang findes i Befolkningsstatistiken, naar Brudgommens og Brudens Aldre stilles i Forhold til hinanden, jfr. omstaaende Tabel I, medens f. Eks. en Tabel med enkelt Indgang kun angiver Brudgommenes Antal i forskellige Aldersklasser, jfr. Tabel II. Man kan sammenligne de to Tabelformer henholdsvis med en grafisk Fremstilling i Rummet og i Planen; i sidstnævnte kan der kun afsættes ud ad 2 Akser, man kan kun fremstille x og y's indbyrdes Afhængighed, i først-

*) Den væsentligste Litteratur angaaende dette Emne: Bowley: Elements of Statistics, London 1901, p. 316—328. Davenport: Statistical methods with special reference to biological variation, New York 1899. Duncker: Die Methode der Variationsstatistik, Leipzig 1899. Norton: Statistical studies in the New York money-market, New York 1902. Pearson: Mathematical contributions to the theory of evolution i Philos. Transact. Roy. Soc. London, 1894, 1895 og 1896, samt talrige andre Afhandlinger i samme. Yule: On the theory of correlation i Journal of the Roy. Stat. Society, 1897.

nævnte kan der afsættes ud ad 3 Akser, her angives x, y og z's indbyrdes Afhængighed. I Tabel II angiver x saaledes Brudgommenes Aldre og y Brudgommenes Antal, i Tabel I angives desuden Brudenes Aldre ved z.

Tabel I.*)

	Brudgommens Alder											
Brudens Alder	Under 25	25—30	30—35	35—40	40—45	45—50	50—55	55—60	60—65	65—70	over 70	Tilsammen
Under 20	2630	2071	602	168	46	21	12	5	»	»	»	5555
20—25..	11055	13426	4515	1344	392	149	75	33	15	6	2	31012
25—30..	4795	10158	5009	1776	674	299	148	57	24	14	8	22962
30—35..	1042	2732	2637	1403	757	345	180	91	40	19	8	9254
35—40..	253	668	814	791	545	339	230	117	46	38	10	3851
40—45..	45	182	266	270	376	278	207	129	63	38	18	1872
45—50..	14	48	70	106	163	215	169	111	70	43	18	1027
50—55..	4	13	26	26	42	69	98	86	63	30	16	473
55—60..	2	2	7	5	9	18	46	53	38	29	15	224
60—65..	1	»	»	»	6	5	8	22	23	22	12	99
65—70..	»	1	1	»	2	3	2	6	3	10	7	35
over 70.	»	»	»	1	»	»	»	»	2	3	11	17
Tilsammen .	19841	29301	13947	5890	3012	1741	1175	710	387	252	125	76381

Tabel II.*).

Brudgommens Alder:	Antal:
Under 25 Aar	19841
25—30	29301
30—35	13947
35—40	5890
40—45	3012
45—50	1741
50—55	1175
55—60	710
60—65	387
65—70	252
over 70	125

*) Statistisk Tabelværk, 4de Række, Litra A, Nr. 9.

Vi vil i det følgende særlig rette vor Opmærksomhed paa Tabellen med dobbelt Indgang. Tabel I angiver Mandens og Kvindens Alder ved Ægteskabets Stiftelse; paa Grundlag af denne har Statens Statistiske Bureau*) udarbejdet følgende to Tabeller:

Tabel III.

Naar Mandens Alder ved Ægteskabets Stiftelse er:	gifter han sig med Kvinder i Alderen:	Antallet af saadanne Ægteskaber er:
22.5	23.9	19841
27.5	25.4	29301
32.5	27.5	13947
37.5	29.8	5890
42.5	33.2	3012
47.5	36.2	1741
52.5	38.9	1175
57.5	42.0	710
62.5	44.7	387
67.5	46.5	252
72.5	49.9	125

Tabel IV.

Naar Kvindens Alder ved Ægteskabets Stiftelse er:	gifter hun sig med Mænd i Alderen:	Antallet af saadanne Ægteskaber er:
18.0	26.2	5555
22.5	27.3	31012
27.5	29.3	22962
32.5	32.9	9254
37.5	37.5	3851
42.5	42.8	1872
47.5	47.6	1027
52.5	51.9	473
57.5	56.4	224
62.5	60.5	99
67.5	60.2	35
72.5	68.4	17

*) l. c.

I Tabel III betragtes Mændenes Aldre som supponerede Variable og Kvindernes Aldre som underordnede Variable, i Tabel IV omvendt. De to Tabeller ere saaledes væsentligt forskellige; den første udviser, hvilken Alder den Kvinde i Gennemsnit har, som en Mand af en given Alder gifter sig med, den anden udviser, hvilken Alder den Mand i Gennemsnit har, som en Kvinde af en given Alder gifter sig med.

Vor Undersøgelse gaar nu først ud paa at finde en Ligning for hvert af de to nævnte Forhold. Mandens Alder betegnes i Tabel III og IV henholdsvis ved x_1 og x_2, Kvindens ved y_1 og y_2 og Ægteskabernes Antal ved h_1 og h_2. I den første Tabel spørger man om, hvilken Værdi y har, naar x kendes, i den anden maa Spørgsmaalet være, hvilken Værdi x har, naar y kendes. Med andre Ord, vi maa hos den første Ligning kræve Formen $y = f(x)$, hos den anden Formen $x = f(y)$.

Vi antager nu, at den første Ligning har Formen

$$y_1 = a \cdot x_1$$

hvor a's Værdi findes ved Hjælp af de mindste Kvadraters Metode som $\frac{\Sigma x_1 \cdot y_1 \cdot h_1}{\Sigma x_1^2 h_1}$, og den anden Ligning Formen $$x_2 = b \cdot y_2.$$
hvor b's Værdi er lig $\frac{\Sigma x_2 \cdot y_2 \cdot h_2}{\Sigma y_2^2 \cdot h_2}$.*)

Samtidig ved man, at Middeltallene

$$M_{x_1} = \frac{\Sigma x_1 h_1}{\Sigma h_1},\ M_{y_1} = \frac{\Sigma y_1 h_1}{\Sigma h_1},\ M_{x_2} = \frac{\Sigma x_2 h_2}{\Sigma h_2}$$
$$\text{og } M_{y_2} = \frac{\Sigma y_2 h_2}{\Sigma h_2}.$$

*) Indsætter man Tallene fra Tabel III og IV faas: $\Sigma h_1 x_1^2 = 73489231^1/_4$, $\Sigma h_1 x_1 y_1 = \Sigma h_2 x_2 y_2 = 63865692^1/_2$ og $\Sigma h_2 y_2^2 = 58332983$ samt $a = 0.8691$ og $b = 1.0930$.

Da Tallene ere beregnede paa Grundlag af en Tabel med dobbelt Indgang, fremkommer der en Række Ligninger mellem h'erne, x'erne og y'erne, nemlig

$$\Sigma h_1 = \Sigma h_2,\ \Sigma x_1 h_1 = \Sigma x_2 h_2,\ \Sigma y_1 h_1 = \Sigma y_2 h_2$$

og $\Sigma x_1 y_1 h_1 = \Sigma x_2 y_2 h_2$, hvoraf igen faas

$$M_{x_1} = M_{x_2} \text{ og } M_{y_1} = M_{y_2},$$

derimod er $\Sigma x_1^2 \gtrless \Sigma x_2^2,\ \Sigma y_1^2 \lessgtr \Sigma y_2^2,$

$$\Sigma x_1^2 h_1 \lessgtr \Sigma x_2^2 h_2 \text{ og } \Sigma y_1^2 h_1 \gtrless \Sigma y_2^2 h_2.$$

Ovenstaaende Tabel I med dobbelt Indgang kan ogsaa betragtes som en Tabel med enkelt Indgang, nemlig saaledes:

Brudgommens Alder (x)	Brudens Alder (y)
$22^1/_2$	18
$22^1/_2$	18
$22^1/_2$	18
o. s. v. ialt 2630 Gange	
$22^1/_2$	$22^1/_2$
$22^1/_2$	$22^1/_2$
$22^1/_2$	$22^1/_2$
o. s. v. ialt 11055 Gange	
o. s. v.	

I denne Tabel kan vi dels lade Brudgommens Alder, dels Brudens Alder være den supponerede Variable. Med andre Ord, vi kan danne to Ligninger:

$$y = a \,.\, x \text{ og}$$
$$x = b \,.\, y,$$

hvor a's og b's Værdi ligesom tidligere er henholdsvis

$$\frac{\Sigma x y}{\Sigma x^2} \text{ og } \frac{\Sigma x y}{\Sigma y^2}.$$

Disse Værdier af a og b ere identiske med de tidligere fundne, idet

$$\Sigma x y = \Sigma h_1 \,.\, x_1 \,.\, y_1 = \Sigma h_2 \,.\, x_2 \,.\, y_2$$
$$\Sigma x^2 = \Sigma h_1 \,.\, x_1^2 \text{ og}$$
$$\Sigma y^2 = \Sigma h_2 \,.\, y_2^2.$$

Resultatet af denne Undersøgelse bliver altsaa, at selv om man kun har en Tabel med enkelt Indgang, er det dog muligt at danne de to Ligninger ved henholdsvis at gaa ud fra, at den ene eller den anden af de iagttagne Værdier er den supponerede; dette spiller en væsentlig Rolle, naar Tabellen med enkelt Indgang kun har faa Iagttagelser af samme Art, f. Eks.:

Brudgommens Alder		Brudens Alder
$22^1/_2$		18
$22^1/_2$		$22^1/_2$
$22^1/_2$		$27^1/_2$
	o. s. v.	
$27^1/_2$		18
$27^1/_2$		$22^1/_2$
$27^1/_2$		$27^1/_2$
	o. s. v.	

Af Hensigtsmæssigheds Grunde omskrives nu de forskellige Summationer paa følgende Maade:

$$\Sigma x_1^2 h_1 = \mu_1^2 . n$$
$$\Sigma y_1^2 h_1 = \mu_2^2 . n$$
og $$\Sigma x_1 y_1 h_1 = \mu_1 . \mu_2 . r_1 . n$$
samt $$\Sigma x_2^2 h_2 = \varrho_1^2 . n$$
$$\Sigma y_2^2 h_2 = \varrho_2^2 . n$$
og $$\Sigma x_2 y_2 h_2 = \varrho_1 \varrho_2 . r_2 . n,$$

hvoraf igen faas

$$a = \frac{\mu_2}{\mu_1} . r_1 \text{ og } b = \frac{\varrho_1}{\varrho_2} . r_2.$$

Hvad særlig Tabellen med enkelt Indgang angaar, saa faas

$$\Sigma x^2 = \Sigma x_1^2 h_1 = \mu_1^2 . n$$
$$\Sigma y^2 = \Sigma y_2^2 . h_2 = \varrho_2^2 . n$$
$$\Sigma xy = \mu_1 . \varrho_2 . n . r,$$

hvoraf igen

$$a = \frac{\varrho_2}{\mu_1} \cdot r \text{ og } b = \frac{\mu_1}{\varrho_2} \cdot r.$$

I sidste Tilfælde kan man udtrykke a og b ved det samme r, noget der ikke er Tilfældet i Tabellen med dobbelt Indgang, saalænge $\frac{\varrho_2}{\mu_1} \gtrless \frac{\mu_2}{\varrho_1}$. Vi ved nemlig, at $\Sigma x_1 y_1 h_1 = \Sigma x_2 y_2 h_2$, altsaa $\mu_1 . \mu_2 . r_1 . n =$ $\varrho_1 . \varrho_2 . r_2 . n_1$ og $r_1 = r_2 . \frac{\varrho_1 \varrho_2}{\mu_1 \mu_2}$; d. v. s. $r_1 = r_2$, naar $\frac{\varrho_1 \varrho_2}{\mu_1 . \mu_2} = 1$. Men den sidstnævnte Ligning behøver ikke at eksistere, da ϱ_1 og μ_2 dannes af beregnede Gennemsnitsværdier og ikke som ϱ_2 og μ_1 af Iagttagelsesværdier.

I Praksis gælder det mere om de enkelte Iagttagelsesværdiers Afvigelse fra deres Gennemsnit end om selve Iagttagelsesværdierne; Spørgsmaalet bliver, hvor stor Afvigelsen fra y'ernes Gennemsnit vil være ved en given Afvigelse fra x'ernes Gennemsnit. Har man en Tabel over Forbruget af en eller anden Vare samt denne Vares Pris i en Række af Aar, kan man spørge om, hvor meget Forbruget vil være over det sædvanlige Gennemsnit, naar Prisen er saa og saa mange Procent under sit Gennemsnit.

I saa Tilfælde vil Udviklingen dog være fuldstændig identisk med den ovenfor givne, kun at x og y saa betegner Afvigelserne for Gennemsnittet, hvoraf igen følger, at μ_1 repræsenterer Middelfejlen paa x'erne, ϱ_2 Middelfejlen paa y'erne, idet vi ved, at Middelfejlens Kvadrat er lig Summen af Afvigelsernes Kvadrater divideret med Antallet, og $\mu_1^2 = \frac{\Sigma x^2}{n}$ og $\varrho_2^2 = \frac{\Sigma y^2}{n}$, hvor Σx^2 og Σy^2 netop er Summen af Afvigelsernes Kvadrater.

Vor Opmærksomhed maa særlig knyttes til Størrelsen r i de Tilfælde, hvor man betragter Afvigelsen fra Gennemsnittet. Denne Størrelse r er først udledet af den franske Matematiker Brawais 1846, senere i 80'erne af Galton. Størrelsen er et Maal for Korrelationen mellem to Iagttagelsesrækker, et Maal for den gensidige Afhængighed, et Maal for Aarsagssammenhængen.

Værdien af r er ifølge det foregaaende lig

$$\frac{\Sigma x y}{n \cdot \sqrt{\frac{\Sigma x^2}{n}} \cdot \sqrt{\frac{\Sigma y^2}{n}}} = \frac{\Sigma x y}{n \cdot \sigma_1 \cdot \sigma_2}$$

hvor σ_1 og σ_2 er Middelfejlene, knyttede henholdsvis til x og y.

r kan aldrig være større end $+ 1$ eller mindre end $- 1$. Man ved nemlig, at

$$\Sigma (y - a \cdot x)^2 = \Sigma y^2 + a^2 \cdot \Sigma x^2 - 2a \cdot \Sigma x y =$$
$$n \, \sigma_2^2 + a^2 \cdot n \cdot \sigma_1^2 - 2a \cdot n \cdot r \cdot \sigma_1 \cdot \sigma_2 =$$
$$n \cdot \sigma_2^2 + \frac{\sigma_2^2}{\sigma_1^2} r^2 \cdot n \cdot \sigma_1^2 - 2 \cdot \frac{\sigma_2}{\sigma_1} \cdot n \cdot r^2 \cdot \sigma_1 \, \sigma_2 =$$
$$n \cdot \sigma_2^2 (1 - r^2),$$

denne Størrelse maa qua Kvadratsum altid være positiv, altsaa $(1 - r^2) > 0$.

Dersom $r = + 1$, maa hver sammenhørende Værdisæt af x og y fuldstændig nøjagtigt tilfredsstille Ligningen; eller med andre Ord Korrelationen er fuldstændig, idet en Forandring i x netop frembringer den ventede Forandring i y, x er ene Aarsag til Forandring i y. At r er positiv, vil kun sige, at de to Iagttagelsesrækker varierer i samme Retning, at r er negativ, at de varierer omvendt.

Dersom $r = \frac{\Sigma x y}{n \cdot \sigma_1 \cdot \sigma_2} = 0$, maa $\Sigma x y = 0$, d.v.s. Produktsummerne af de positive og de negative Afvigelser maa fuldstændig ophæve hinanden, noget der er Tilfældet, hvis x og y er tilfældig kombinerede, hvad der igen vil sige, at Korrelationen er Nul.

I Almindelighed vil man faa r liggende mellem 0 og $+1$; i saa Tilfælde er der delvis Korrelation, det vil sige der findes andre Aarsager end x, der paavirker y; jo nærmere r er ved $+1$, des bedre maa Korrelationen være og omvendt.*)

Før vi gaar videre, skal vi gennemgaa et Eksempel taget fra Bowley (l. c.).

I nedenstaaende Tabel angiver Talrækken X den aarlige Ægteskabshyppighed, Talrækken Y Hvedepriserne, begge for Aarene 1875—94. Det gælder om at vise, hvor stor Korrelationen er mellem disse to økonomiske Fænomener. Man danner i det Øjemed Talrækkerne x og y; den første angiver Afvigelsen mellem X og Gennemsnittet $\frac{\Sigma X}{n} = 15.17$, den anden Afvigelsen mellem Y og Gennemsnittet $\frac{\Sigma Y}{n} = 37.10$.

Endelig dannes en 5te Kolonne xy simpelthen ved Multiplikation af Kolonne x og y. Summen af denne Kolonne $\Sigma x y = 627$.

*) Til hver af de ovenfor beregnede Størrelser knytter sig en Middelfejl; saaledes har M_x Middelfejlen $\frac{\sigma_x}{\sqrt{n}}$, σ Middelfejlen $\frac{\sigma}{\sqrt{2n}}$, r Middelfejlen $\frac{(1 - r^2)}{\sqrt{n}}$ og $r \cdot \frac{\sigma_1}{\sigma_2}$ Middelfejlen $\frac{\sigma_1}{\sigma_2} \cdot \sqrt{\frac{1-r^2}{n}}$ (Pearson og Filon i Phil. Trans. 1898.)

Endvidere beregnes $\sigma_1 = \sqrt{\frac{\Sigma x^2}{n}}$ og $\sigma_2 = \sqrt{\frac{\Sigma y^2}{n}}$ til henholdsvis 0.651 og 102.

Nu har vi alle de nødvendige Faktorer til Bestemmelsen af

$$r = \frac{627}{20 \,.\, 102 \,.\, 0.651} = +0.47,$$

idet Antallet af Aar (n) er lig 20.

Tabel V.

Aar	Ægteskabshyppighed X	x	Hvedepris Y sh.	d.	y	xy
1875	16.7	+ 1.53	45	2	+ 88	+135
1876	16.5	+ 1.33	46	2	+100	+133
1877	15.7	+ .53	56	9	+227	+120
1878	15.2	+ .03	46	5	+103	+ 3
1879	14.4	— .77	43	10	+ 72	— 55
1880	14.9	— .27	44	4	+ 78	— 21
1881	15.1	— .07	45	4	+ 90	— 6
1882	15.5	+ .33	45	1	+ 87	+ 29
1883	15.5	+ .33	41	7	+ 45	+ 15
1884	15.1	— .17	35	8	— 26	+ 4
1885	14.5	— .67	32	10	— 60	+ 40
1886	14.2	— .97	31	0	— 82	+ 80
1887	14.4	— .77	32	6	— 64	+ 49
1888	14.4	— .77	31	10	— 72	+ 55
1889	15.0	— .17	29	9	— 97	+ 16
1890	15.5	+ .33	31	11	— 72	— 23
1891	15.6	+ .43	37	0	— 10	— 4
1892	15.4	+ .23	30	3	— 91	— 21
1893	14.7	— .47	26	4	—138	+ 65
1894	15.1	— .07	22	10	—180	+ 13

Vi har talt om Korrelation mellem Afvigelserne fra Gennemsnittet; man kunde nu spørge, om Begrebet r, der ogsaa findes ved Bearbejdelse af selve Iagttagelserne, ligeledes giver et Udtryk for Korrelationen mellem disse. Hertil maa dog delvis svares nej. Er Ligningen mellem Afvigelserne

$$y = ax,$$

bliver Ligningen mellem Iagttagelserne (X og Y)

$$Y - M_y = a(X - M_x) \text{ og}$$
$$Y = aX + K.$$

Udtrykket $\Sigma(Y - aX - K)^2 = v^2(1 - r^2)$ maa altid være positivt, altsaa maa r ligge mellem ± 1, men da $\Sigma\, YX$ aldrig kan blive Nul, saalænge Y og X begge ere positive Størrelser, kan $r = \frac{\Sigma\, YX}{\sqrt{\Sigma\, Y^2 \cdot \Sigma\, X^2}}$ aldrig blive Nul.

Med andre Ord naar Resultatet af en Undersøgelse giver $r = \pm 1$, kan man hævde fuldstændig Korrelation; for alle andre Værdier af r kan man derimod intet udsige om Korrelationen.

Ved en økonomisk-statistisk Undersøgelse af denne Art er der endnu en Række Bemærkninger at gøre.

Har man f. Eks. Ægteskabshyppigheden og Importen pr. Individ en Række af Aar og spørger om Korrelationen herimellem, saa er det bedst først at underkaste Iagttagelserne en særlig Behandling. Bevægelsen i Importen saavel som Bevægelsen i Ægteskabshyppigheden fra Aar til Aar skyldes to forskellige Grupper af Aarsager: en Gruppe af Aarsager, der virker i en enkelt bestemt Retning, og en anden Gruppe, der virker snart i en, snart i en anden Retning. Med Samfundsudviklingen stiger Importen paa Grund af en stadig virkende (ganske vist snart stærkere snart svagere) Række Aarsager, medens Svingningerne fra Aar til Aar skyldes en helt anden Række Aarsager; ligeledes ved Ægteskabshyppigheden.

Det gælder her at eliminere den Bevægelse, som Samfundsudviklingen (eller hvad man nu vil kalde det) foraarsager, saa at man kun har at gøre med de særlige Aarsager fra Aar til Aar, og saa besvare Spørgsmaalet, hvorvidt disse særlige Aarsager, der paavirker x eller y, udelukkende er y henholdsvis x eller andre.

En saadan Eliminering kan foretages paa mange forskellige Maader. Norton (l. c.) antager, at Samfundsudviklingen lader Iagttagelserne vokse efter den almindelige Renteformel $a(1 + v)^t$. Man maa indvende mod denne Fremgangsmaade, at dens Forudsætning konstant v ubetinget er ukorrekt; Samfundsudviklingen er snart svagere, snart stærkere, kun indenfor et begrænset Tidsrum er Forudsætningen gyldig. Hooker*) foreslaar at erstatte hver Iagttagelse med Gennemsnittet af denne, de to foregaaende og de to efterfølgende; denne Fremgangsmaade er ganske vist ret vilkaarlig, men dog langt at foretrække for Nortons. Man kan efter en saadan Bearbejdelse benytte $r = \frac{\Sigma xy}{n \cdot \sigma_1 \cdot \sigma_2}$, idet x og y betyder Afvigelserne mellem de iagttagne og de beregnede Værdier, og $\Sigma xy = 0$ omtrent svarer til $r = 0$.

Ret naturlig synes endnu en lille Ændring; i Stedet for at regne med Afvigelsen mellem den iagttagne og den beregnede Værdi, saa at regne med denne Afvigelse sat i Forhold til den iagttagne Værdi.

Hele vor Undersøgelse angaaende Korrelation hviler paa, at den antagne Ligning $y = ax$ er nogenlunde rigtig; dersom nemlig Ligningen mellem y og x

---

*) Correlation of the marriage-rate with trade i Journal of the Roy. Stat. Society 1901.

er en anden, kan r, naar man benytter Ligningen $y = ax$, aldrig blive $+ 1$, om end den som oftest ikke vil afvige meget derfra.*)

Har Kurven saaledes den ret almindelige Form $y = \frac{a}{x^k}$, kan man vanskelig uden grov Unøjagtighed tvinge den til at være $y = ax$. Bedst er det i saa Tilfælde — efter at have benyttet Hookers Metode og

---

*) Ligningen mellem Afvigelserne kan skrives under Formen

$$y = ax + bx^3 + cx^5 + \dots + vx^{2n-1} + \dots$$

hvor a, b, c .... maa have samme Fortegn, da man samtidig skal have opfyldt Betingelsen $\Sigma x = 0$ og $\Sigma y = 0$.

Korrelationen r bliver saaledes, naar Ligningen har Formen $y = ax + bx^3$ (istedetfor Formen $y = ax$) lig

$$\frac{\Sigma xy}{\sqrt{\Sigma x^2} \cdot \sqrt{\Sigma y^2}} = \frac{\Sigma (ax^2 + bx^4)}{\sqrt{\Sigma x^2} \cdot \sqrt{\Sigma (ax + bx^3)^2}}$$

$$= \sqrt{\frac{a^2 (\Sigma x^2)^2 + b^2 (\Sigma x^4)^2 + 2ab\, \Sigma x^2\, \Sigma x^4}{a^2 (\Sigma x^2)^2 + b^2\, \Sigma x^2\, \Sigma x^6 + 2ab\, \Sigma x^2\, \Sigma x^4}}$$

Her er $(\Sigma x^4)^2 < \Sigma x^2\, \Sigma x^6$, men Forholdet mellem de to Størrelser er nær ved 1.

Afvigelserne fordeler sig i store Træk efter Eksponentialformlen, der groft kan udtrykkes ved følgende Tal:

25%	af Tilf.	har i	Gnst. en	Afvig.	af	0.2	Gange	Middelfejlen
25%	—	—	—			0.5	—	—
25%	—	—	—			1.0	—	—
25%	—	—	—			1.8	—	—

Med andre Ord vi kan tilnærmelsesvis sætte Værdierne af x lig 2, 5, 10 og 18; i saa Fald faas

$$(\Sigma x^4)^2 = c. \quad 13125 \text{ Millioner}$$

$$\Sigma x^6 \cdot \Sigma x^2 = c. \quad 15750 \text{ Millioner}$$

Forholdet mellem disse to Tal er 0.83.

Men er $\frac{\alpha}{\beta} = 0.83$ maa $\frac{a + \varepsilon}{b + \varepsilon} > 0.83$ og fuldstændig Korrelation maa mindst have Værdien $\sqrt{0.83} = 0.91$.

Paa lignende Maade kan man gaa frem overfor en Ligning af Formen $y = ax + bx^3 + cx^5$, hvor man vil finde, at der til fuldstændig Korrelation kræves mindst 0.85, af Formen $y = ax + bx^3 + cx^5 + dx^7$ o. s. v.

fundet Afvigelserne x og y — at danne $\log(100 + x)$ og $\log(100 + y)$ og derefter Korrelationen mellem

$$\log(100 + x) - \frac{\Sigma \log(100 + x)}{n} = \delta \log(100 + x)$$

og $$\log(100 + y) - \frac{\Sigma \log(100 + y)}{n} = \delta \log(100 + y)$$

af Ligningen

$$\delta \log(100 + y) = c \,.\, \delta \log(100 + x).$$

Af denne Ligning kan igen dannes Ligningen

$$\log(100 + y) - \frac{\Sigma \log(100 + y)}{n} = c \log(100 + x)$$
$$- \frac{c \Sigma \log(100 + x)}{n}.$$

Hos Økonomerne er Kriteriet r kun benyttet meget lidet. Man har fastslaaet Korrelation mellem følgende økonomiske Fænomener, naturligvis kun gyldig for et bestemt Sted og til en given Tid.

Økonomisk Fænomen	r	Forfatter
Fattigdom og Fattighjælp	+ 0.34	Yule
Fattigdom og Løn	— 0.66	—
Ægteskabshyppighed og Import	+ 0.79	Hooker
— og Import + Eksport (1861—95)	+ 0.86	—
— og Clearinghouse	+ 0.47	—
— og Hvedepriser (1861—95)	+ 0.09	—
— og Eksport	+ 0.80	—
— og Hvedepriser (1845—64)	— 0.29	Bowley
— og Hvedepriser (1875—94)	+ 0.47	—
— og Import + Eksport (1845—64)	+ 0.01	—
— og Import + Eksport (1875—94)	+ 0.25	—
Deposita og Rente	+ 0.52	Norton

H o o k e r har ovenfor beregnet Korrelationen mellem Ægteskabshyppigheden og Eksporten i d e t t i l s v a r e n d e A a r, men har desuden fundet Korrelation mellem Ægteskabshyppigheden og Eksporten et halvt Aar, et helt Aar o. s. v. tidligere eller senere.

Hooker faar følgende Værdier af r for Eksporten $-1\frac{1}{2}$, -1, $-\frac{1}{2}$, 0, $+\frac{1}{2}$, $+1$ Aar efter Ægteskabsaaret: $+0.58$, $+0.78$, $+0.86$, $+0.80$, $+0.61$, $+0.33$.

Sammenlignes Hookers og Bowleys r for de samme Fænomener, saa giver Hookers langt det bedste Resultat; Grunden hertil maa søges i den ovenfor nævnte Bearbejdelse, som Hooker underkaster Tallene.

N o r t o n har beregnet Korrelationen mellem Deposita og Rente paa Grundlag af 780 Erfaringer; han deler disse Erfaringer i to Dele, saaledes at den ene Gruppe indeholder de Erfaringer, hvor Renten er over 4 %, den anden de Erfaringer, hvor Renten er under 4 %; i disse Tilfælde faar Norton henholdsvis 0.59 og 0.60 som r's Værdi. Korrelationen bliver saaledes bedre efter end før Delingen; Grunden hertil ligger simpelthen i, at Kurven $y = ax$ bedre tilfredsstiller enkelte Dele af Kurven end hele Kurven.

G a l t o n*) var som før nævnt efter Brawais den første, der benyttede Begrebet Korrelation; hans Udvikling er dog ikke fuldstændig lig med den ovenfor givne, idet r sættes lig

$$\frac{\frac{\varepsilon_y}{\varepsilon_x} \cdot \Sigma \frac{v_x}{v_y}}{n},$$

hvor v_x og v_y, idet den supponerede Variable kaldes x

*) Natural inheritance, London 1889.

og den underordnede Variable y, henholdsvis er lig $x - \frac{\Sigma x h}{\Sigma h}$ og $y - \frac{\Sigma y h}{\Sigma h}$, samt ε_x og ε_y henholdsvis lig

$$\sqrt{\frac{\Sigma h v_x^2}{\Sigma h}} \text{ og } \sqrt{\frac{\Sigma h v_y^2}{\Sigma h}};$$

h er Hyppighedskoefficienten og n Antallet af Variable.

Da nu saavel x som y kan betragtes som den supponerede — vi har jo ligesom ovenfor at gøre med en Tabel med dobbelt Indgang — faas to Værdier for r, der ligesom tidligere ved Korrelationen mellem de absolute Iagttagelsesværdier ikke er fuldstændig identiske.

Foruden Galton har Fechner*) angivet en — ganske vist ret overfladisk — Metode til Beregning af »Abhängigkeitsverhältnisse«. Metoden bestaar simpelthen i en Opgørelse af, i hvormange Tilfælde de to Fænomener bevæger sig i samme Retning, begge tiltager (+), og i hvormange Tilfælde, det ene Fænomen tiltager, det andet aftager (÷). Er Antallet af Plusser lig a og Antallet af Minusser lig b, kan man, naar enten a eller b er Nul, hævde, at der er fuldstændig Korrelation, og — paastaar Fechner — ingen Korrelation, naar $2a = b$. Delvis Korrelation maa ligge herimellem, maalt ved

$$r = \frac{2a - b}{2(a + b)}.$$

Fechner anfører foruden denne en anden Formel, der ikke har den væsentlige Fejl at give et forskelligt Resultat, eftersom man betegner Plussernes Antal ved a eller b.

Korrelation er efter denne Formel Nul, naar $a = b$, og delvis naar

$$r = \frac{a - b}{a + b}.$$

*) Kollektivmasslehre, Leipzig 1897.