經濟叢書

統計學原理

趙文銳譯

百里先生教正
文銳謹贈

共學社

1923

原序

曩昔一九〇七年，予在牛津大學講演時，嘗謂最新之統計學，其主要之原理，試驗之結果，學術之名詞，較通常之所教授者，可有更簡單表明之方法。幷指導以革新之大綱，進行之規畫，而示方法之當如何。末言深望有勝任之教員，循此程序，悉心研究，以實現通俗教授之方法。蓋世多有熱心統計學之人，徒以無專門之研究，而缺算學之智識，即統計調查之結果，關於彼等科學範圍內者，亦苦理解之無由。苟有通俗之書出而濟其急，其有裨統計學智識之普及者，決非淺鮮。倍林愛爾賓登君(W. Palin Elderton)與其妹愛賽爾愛爾賓登女士(Ethel M. Elderton)，願以合著此通俗之統計學書爲己任，聞之甚喜，喜得人也。其結果有「統計學原理」(Primer of statistics)之成，不久將見最新統計學之最新智識，普及於有教育之人；予以是書卜之矣。

一九〇九年九月佛蘭雪斯高爾登(Sir Francis Galton)序

譯序

原書在統計學上之位置與價值，已見於原序。雖出版於一九〇九年，即在今日，美國大學，仍用爲教本。蓋英文統計學書中，說理最新而據例最淺近者，首推此書。國人科學之智識，甚感缺乏，而統計學之智識爲尤甚。夫以科學昌明之英國，有待乎統計學智識之普及，而如此通俗之書需要若斯其急也，况中國乎？爰譯之以餉國人。

原書力求簡易，譯書更甚。刪蕪去冗，自不待言，即學理之嫌過深，而爲本書之所不必有者，亦淘汰之。且於正文，參以己意，復加案語，以期明瞭。惟括弧內之注解，不冠以「案」字者，爲原書所本有，不敢掠美，附記於此。

民國十一年四月趙文銳識

統計學原理

目錄

統計學原理

英國 倍林 愛賽爾 愛爾竇登合著　　嵊縣趙文銳譯

第一章　個體 Variates 與中位數 Medians

試散步公園，自一樹採集許多葉片，比較之，則見有長短之不同；度量之，則見最長與最短者大相懸殊。不特樹葉呈若斯之現象，卽採集他種物品而觀察之亦然。物之大小，無定也如此，設有人問以某種樹葉某種堅果或某種貝殼之長度，當爲幾何，恐不能答覆。雖可曰此類天然物，以所遭各種境遇之不同，終不免有參差之不齊。但此爲勉强之解釋，不足爲完滿之答覆。然則完滿之答覆，果可得乎？此問題也，請作種種之試驗以解決之。

以何物爲試驗品，可不甚注意，苟其物易移動而易度量足已。今先以貝殼 Shells 爲試驗品。使散播一袋之貝殼，則見紊亂落下，毫無秩序。欲整齊排列之，或以長短，或以寬狹，或以輕重爲

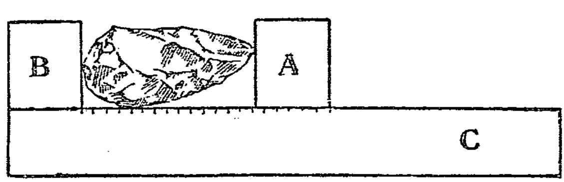
標準，而定其先後之位置，俱無不可。試先度量其長短。度量之簡便器具，有如第一圖，C為長尺，B為固定之木板，A為活動之木板。以試驗品置於長尺之上，便其一端接觸B木板。移動A木板之地位，使其接觸他一端，乃記明其長度。

試取為數不多不少之貝殼五十九個，一一如上法度量之，按其長短之秩序，以等距離排置之。如此有條不紊之排置，統計學家名之曰「整列」。

Array　苟照貝殼各個體之長短，而一一畫直線於底線上，以代表之，則直線全體之整列，有如第二圖。若以一線聯接各直線之上端，則其結果，似足表明許多之貝殼。按秩序排列之，其間長短不同之變化，卻有一定之統系。

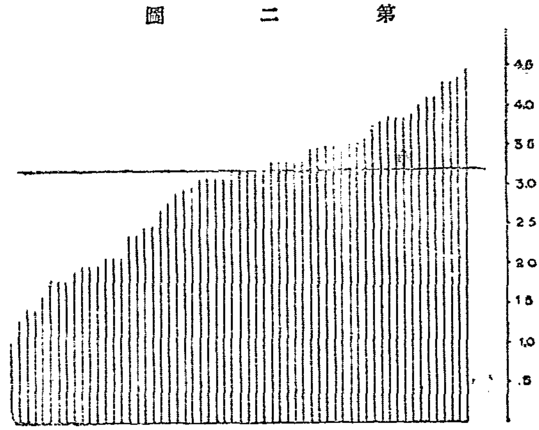
此為發明貝殼長短之變化，確有規則之第一步。但至此而吾人不免起一種之疑問。疑夫前所採集之貝殼，其長短之變化，固有規則，然苟採其

第一圖



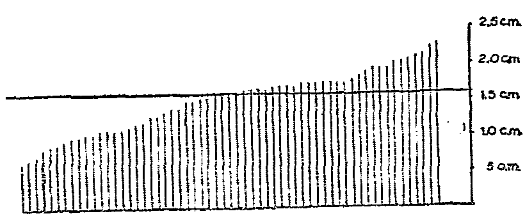
他之貝殼，卽爲同類，而其長短之變化，或無規則，亦未可知。欲試驗之，另採集爲數較少之貝殼三十三個，如前排置之，則作第四圖之「整列」。又將前述之二堆貝殼，各按寬狹之秩序如前排置之。則作第三圖與第五圖之整列。

第二圖



表明五十九個貝殼之長度

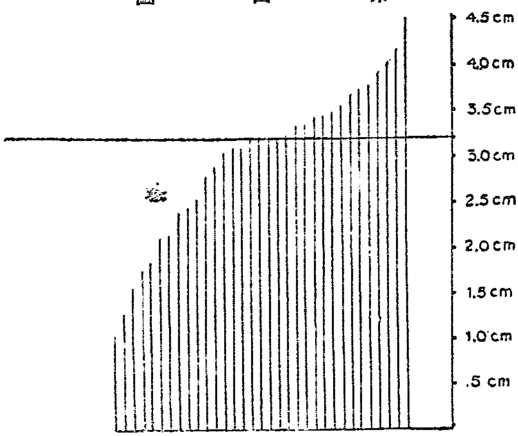
第三圖



表明五十九個貝殼之寬度

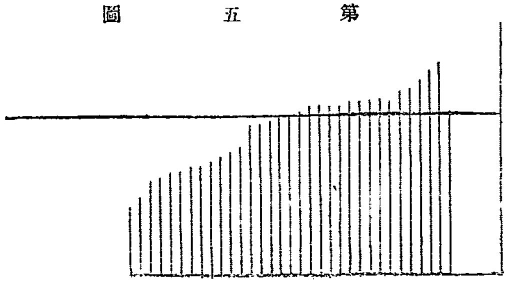
讀者試詳細觀察如此「整列」之四圖，則見有以下相同之點。(一)各圖近中間貝殼之部分皆平坦，而無甚高底。換言之，多數之貝殼或個體 Variates，與居中者之度數幾相等。(二)各圖居中之直線，原所以代表居中之貝殼。在同一秩序之圖中，此居中直線

第四圖



表明三十三個貝殼之長度

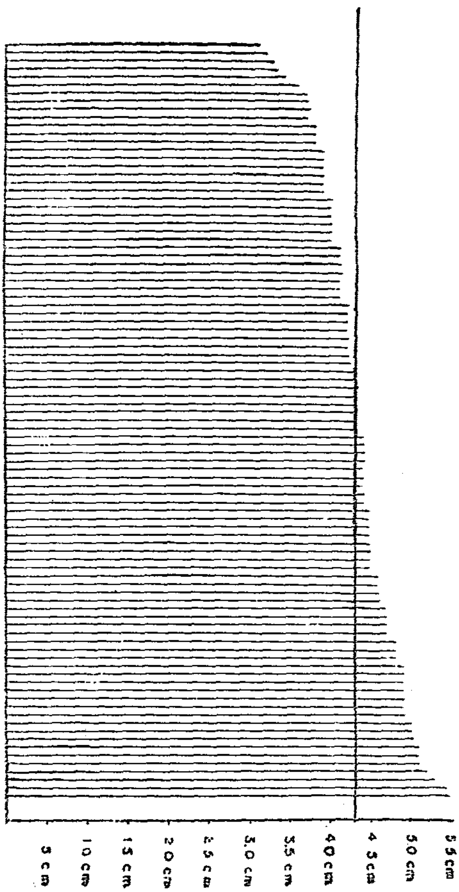
第五圖



表明三十三個貝殼之寬度

與彼居中直線之高度，相差有限。換言之，居中之貝殼，在同一秩序之「整列」中，其度數幾相等。此

第六圖

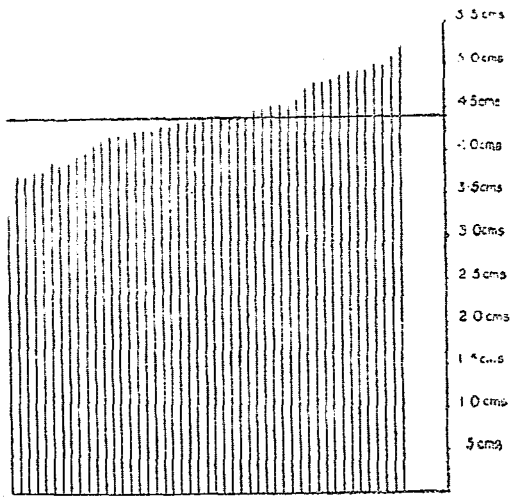


表明一百八十五顆堅果之長度；恐圖過大，僅就相間之個體表明之

居中之貝殼或個體，統計學家名之曰中位數。Median （案，居中之個體，如為單個，即以之為中位數，若為兩個，則取二數平均之，為中位數，）例如第一堆貝殼長度之中位數，為3·23生的米突，第二堆為3·20生的米突，而二者價值之相去無幾。

（三）各圖有同一特殊之形狀，兩端參差，而中部平坦。換言之，即整列中貝殼個體之度數與中位數相較「小變化」Small deviations 多而「大變化」Big deviations 少是已。

第七圖

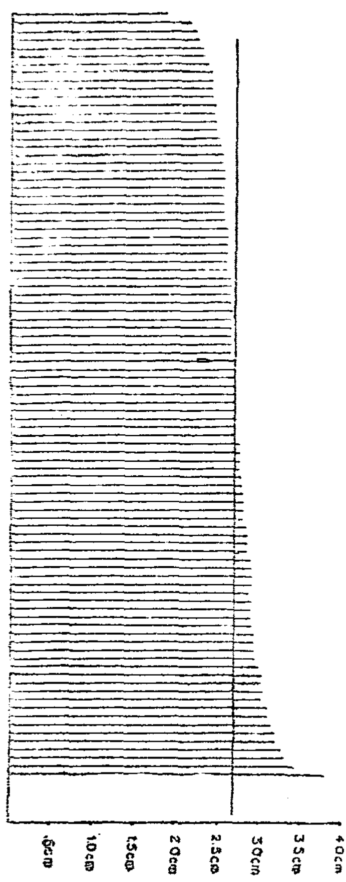


表明四十七顆貝殼之長度

設又採集其他同類之貝殼，當得同一之結果。故得斷言之曰，貝殼有大小適中之軀材，而此

適中軀材之大小，在各種之採集中，常一定不變。然不能推貝殼而謂他種之物品皆然，請以堅果Nuts為試驗品，度量其長短寬狹，而觀其所得之結果如何。先採集一百八十五顆堅果為第一

第八圖



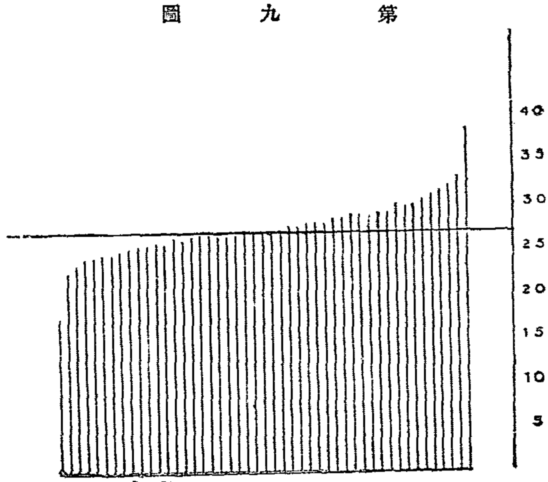
表明一百八十五顆果殼之寬度僅表明相間之個體與第六圖同

堆，繼採集四十七顆為第二堆。先按二堆長短之秩序而各整列之，繼按寬狹之秩序而各整列之，有如第六圖至第九圖之現象。試與貝殼之圖相較，則見二類之圖，大相類似。然則吾人不能不以

為無論何種之「整列」，必有一普通之規則，寓於其間。苟能發明此規則，雖物之大小無定，而吾人可有簡單之推測，得算出一大旨不差之數目矣。

雖然，前之試驗品，若貝殼堅果，既為天然物，或不脫天然之規則，但偶然之遭際則何如。夫擲錢（coin-tossing）之事，五尺童子皆知之，或陰或陽，全屬於偶然，請又以此為試驗品，而觀其所得之結果如何。試以錢十四枚，擲百五十次，每次所起「陽面」之數，先後順序記之如下。

第九圖



表明四十七顆果殼之寬度

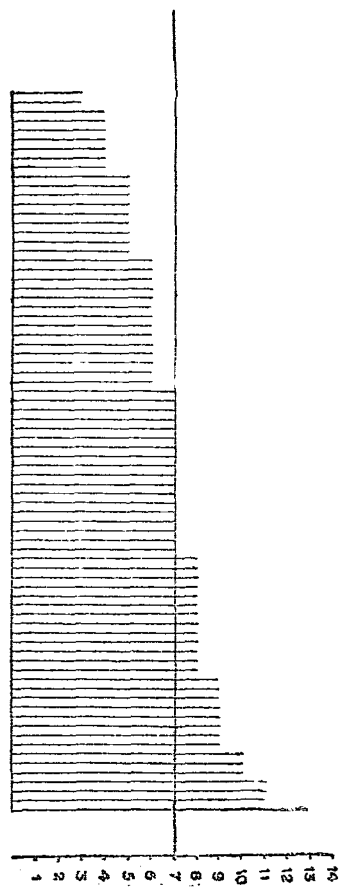
7, 6, 9, 7, 3, 6, 8, 6, 5, 11, 11, 6, 4, 7, 8, 7, 6, 6, 4, 6, 9, 7, 5, 4,

4, 3, 6, 11,8, 8, 8, 7, 6, 6, 6, 5, 10,7, 7, 8, 9, 8, 9, 7, 8, 4, 8, 9, 6, 4,
10, 8, 6, 7, 6, 5, 4, 11,5, 6, 7, 6, 6, 4, 6, 9, 7, 6, 6, 9, 6, 4, 7, 6, 7,
9, 13, 4, 7, 7, 6, 8, 4, 5, 9, 5, 4, 10,8, 7, 9, 8, 5, 7, 9, 7, 11,8, 5, 5, 4,
4, 7, 8, 5, 4, 7, 5, 9, 8, 7, 6, 7, 5, 5, 7, 7, 7, 7, 6, 8, 8, 7, 9, 7, 10,
8, 9, 8, 8, 6, 7, 5, 7, 11,10,7, 4, 8, 7, 5, 10,9, 7, 4, 7, 7, 5, 8, 8, 6.

此爲紊亂無秩序之位置。苟整齊排列之，每畫一直線，以代表每一次所起陽面之數，則呈第十圖之形狀。

觀此圖之曲線，與在前之各圖者類似。其他以偶然之事爲試驗者亦然。然則無論爲天然之物品，偶然之遭際，均得同一之結果。故得斷言之曰(一)迭次採集同類之標本，其所得之中位數，大約相等，故中位數爲度量整列中個體有用的準個。(二)苟作圖以表明試驗之結果，無論所取之標本或爲貝殼，或爲堅果，或爲擲錢，所得之形狀，幾呈一致之趨向。即圖中之曲線，初則速升，繼則平行，終又速升是已。此無他，在整列之中部，與中位數相等之個體不少，在整列之兩端則甚少

第十圖



表明十四枚錢擲陰陽之圖恐過大當略每隔一次之數據

故也。此爲整列圖中曲線極普通之形狀，但亦有不盡然者，試言之於下章。

第三章　四分位數 Quartiles 與平均數 Means

茲爲便於討論計，將前章試驗所得之原理復重述之。苟迭次採集同種物品之個體而度量

之，畫直線以表明其結果，所有之中位數，大約相等，有似一定不變者然，竟可視之為「常價」。Constant or unchanging value 其連接各直線上端所成之曲線，足以形容個體變化之趨勢。欲顯知變化之趨勢，可於中位數之上端畫一平線（如第一章之各圖是）如是則見曲線近中位數之處，較為平坦，而在兩端，則為傾斜。此皆言之於第一章。惟所欲反覆申言者，則前作試驗中之個體，與中位數相較，「大變化」少，「小變化」多。簡言之，多數之個體，與中位數大約相等，是已設「大變化」「小變化」之個體，多少相等，則聯接各直線上端之線，當為直線。今不為直線而為曲線者，則非相等可知。

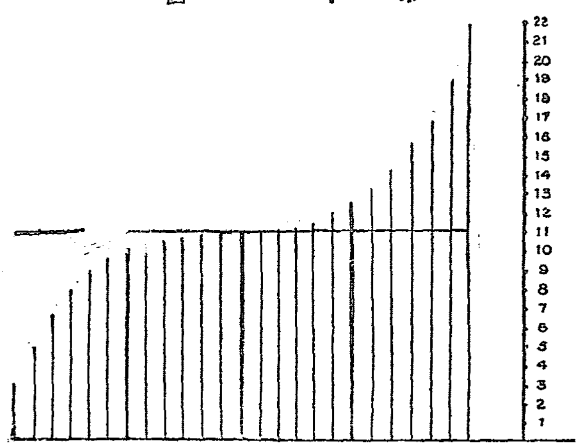
若又進一步而於曲線之情狀，更有所知，——例如能言明曲線在某處升或降，距離經過中位數之平線若何，——則於統計之調查，必有更切實之結果。試以中位數為「常價」，而二分所採集之個體，為前後兩段，使前段所含之個體，小於中位數，後段之個體，大於中位數。復將此兩段各二分之，而記明其中位數之長度。若斯之處置，能使全體分配之情狀，更形顯明。例如前所整列之四十七顆堅果，其中位數為4·38生的米突，前半段之中位數為4·20生的米突，後半段之中位數為

4·77生的米突。如此四分之位，統計學家稱之曰「四分位數」Quartiles 以其均分全體爲四段，而足以表明所調查的事物變化之程度若何也。作四分之方法，觀第十一圖自明。在此圖中，中位數之高度爲11，前四分位數爲$10\frac{1}{5}$，後四分位數爲$12\frac{3}{5}$。

設推廣構造四分之思想，而均分全體爲百分，亦無不可。然苟考察之物品，爲數加多，普通宜改變方法以處置之，於是統計學上有分析更精密之方法，請言之於後。今將所考察者更詳細研究，進而論整列中個體之平均數。Mean

平均數之意義，人人知之。其計算之方法，亦

第十一圖



人人知之。欲知一羣小孩之平均歲數，總加其歲數，而以所有孩童之總數除之卽得。同此理也欲知所採集貝殼長度之平均數，總加其度數，而以所有貝殼之總數除之卽得。

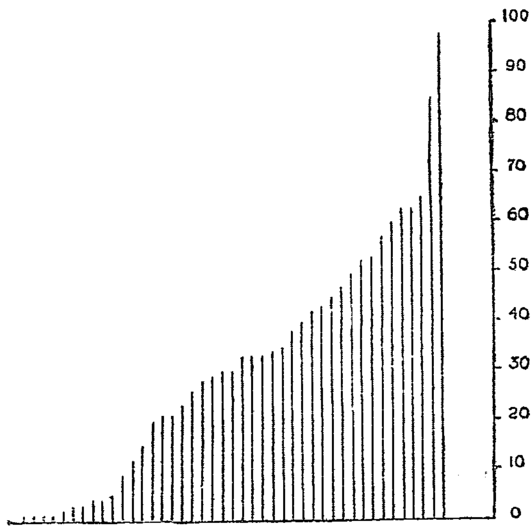
惟平均數與中位數有何區別，此爲吾人所待討論者。試以棍球之點數 Cricket scores 爲例。（案，打棍球爲英國一種之團體運動，其規則甚複雜，恐非讀者易於了解，故本書作爲個人運動觀，如30或30·2點數，卽作爲個人每次打棍球所得之點數爲30或30·2是。雖若斯之解釋，不免失其本來之性質，但不能代之以相當之例，而於本書學理之討論，依然無傷，故亦不足爲病）將譚泉克力夫 Tunnicliffe 在1907年時，所打棍球之點數，通盤計算之，則見平均點數爲30·2，中位點數爲30（見第十二圖）。此二數所表明之意義，卽謂譚泉克力夫在1907年時，打棍球之點數，有時少於30，有時多於30，其平均點數則爲30·2是已。

讀者試以第十二圖，與第一章所作貝殼堅果之諸圖相較，則見二種之形狀，大不相同。前章所作貝殼或堅果之諸圖，其聯接各直線上端之曲線，幾爲平線，以有大小幾等之個體不少也。第十二圖則不然，其聯接各直線上端之曲線，幾自底下之平線起，而甚傾斜，以最初與最後時所得

之點數，大相懸殊也。棍球點數，現如此特殊之狀態者，以棍球每年初出打時，尚少練習，故所得之點數甚底。不特譚皋克力夫如此，即考察其他好手之點數，所得之結果大概相同。如此整列，詳論之於後。茲所論者，僅爲平均數與中位數大小之比較。故將前所舉之各例，計算其平均數與中位數，併將某學校在1897年冬季時一百學生年齡之平均數與中位數，匯列爲第一表。

由上表以觀平均數與中位數，

第十二圖



表明譚皋克力夫在1907年所得之點數

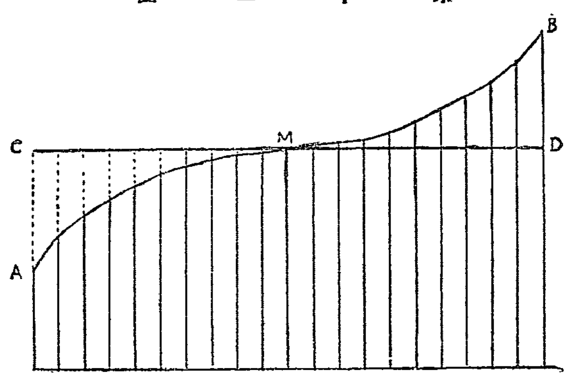
第一表

	平均數	中位數
47堅果(長度)	4·40cm.	4·38cm.
25堅果(長度)	4·24cm.	4·24cm.
33貝殼(長度)	3·20cm.	3·20cm.
33貝殼(寬度)	1·53cm.	1·61cm.
59貝殼(長度)	3·03cm.	3·23cm.
59貝殼(寬度)	1·52cm.	1·60cm.
41貝殼(長度)	3·21cm.	3·21cm.
41貝殼(寬度)	1·58cm.	1·56cm.
100貝殼(長度)	3·10cm.	3·20cm.
100貝殼(寬度)	1·55cm.	1·59cm.
譚臬克力夫(棍球點數)	30·2	30
學生之年齡	十三歲零四月	十三歲零六月

相差有限。然則須有如何條件，而二數適相等，請討論之。第十三圖，Om爲中位數，AmB爲表明個體變化之曲線，CmD爲經過中位數上端之直線，一切構造，與前所作之圖皆同。欲得平均數之價值，總加各個體之度數，而除之以所有個體之總數卽得。在平均數後較大之個體，有在平均數前較小之個體以平衡之。兩旁大小之數，適足相抵，故平均數與中位數相等。而在此例，不特二數相等，且全體之變化，又皆勻稱。Symmetrical 以前後各相當之位數，Corresponding terms 一大一小，適足相抵也。此爲平均數與中位數相等之一例，而又有其他之例。然無論如何，其相等之條件，必mD直線與mB曲線間之部分，等於Cm直線及Am曲線間虛設之部分而後可。

本章所論，可概括之如下。在前章中，吾人已知「整列」中之個體，「大變化」不如「小變化」之爲常見。在本章中吾人知除中位數外，又有其他之價值如四分位數者，可藉以測知近旁個體變化之速度。且知平均數之如何算定，而見其價值與中位數相差不遠。又知平均數與中位數相等時所必須之條件。一言以蔽之，中位數表明平均大小之價值，四分位數表明價值

第十三圖



變化之趨勢。

第三章　次數之分配 Frequency Distributions

前二章所作之試驗，皆就個體而言。然有時所採集之標本，爲數較多，欲逐一指明其長短寬狹，似覺不便，不如就其相間之個體表明之，事簡而功亦相等。若斯所試驗之物品，不啻區別爲各種階級，而所作之試驗，可作爲團體之試驗觀。

惟所宜注意者，苟採集物品，排置之成團體，各團體間之距離宜相等。且此團體與彼團體之間，無論有項與否，當注明其限度，不得省畧之。（讀者請參考下列「例二」第二項「例五」第一項第二項與第十二項，卽明此義）

將堅果與貝殼，依上法排置之，則成如下之統系。

〔例一〕　堅果之長度（以生的米突爲單位）

果數	2	7	28	59	49	33	6	1
長度	2·8—	3·2—	3·6—	4·0—	4·4—	4·8—	5·2—	5·6—

〔例二〕　堅果之寬度(以生的米突爲單位)

果數	2	0	5	22	55	48	.36	10	5	1	1
寬度	1·7—	1·9—	2·1—	2·3—	2·5—	2·7—	2·9—	3·1—	3·3—	3·5—	3·7—

〔例三〕　貝殼之長度(以生的米突爲單位)

殼數	2	5	11	15	25	24	15	3
長度	0·9—	1·4—	1·9—	2·4—	2·9—	3·4—	3·9—	4·4—

〔例四〕　貝殼之寬度(以生的米突爲單位)

殼數	3	9	11	14	23	23	9	6
寬度	0·7—	0·9—	1·1—	1·3—	1·5—	1·7—	1·9—	2·1—

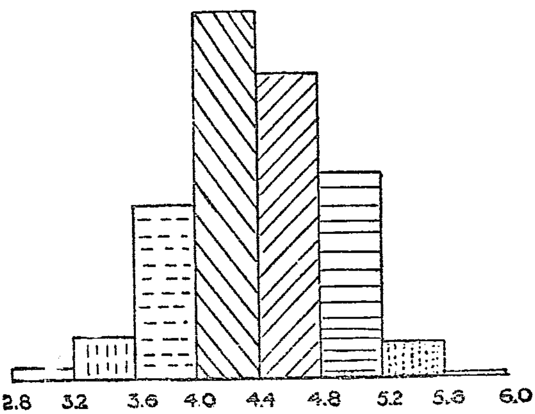
〔例五〕　擲錢(十四枚)

次數	0	0	2	15	17	27	36	25	15	6	6	0	1	0
每次陽面之數	1	2	3	4	5	6	7	8	9	10	11	12	13	14

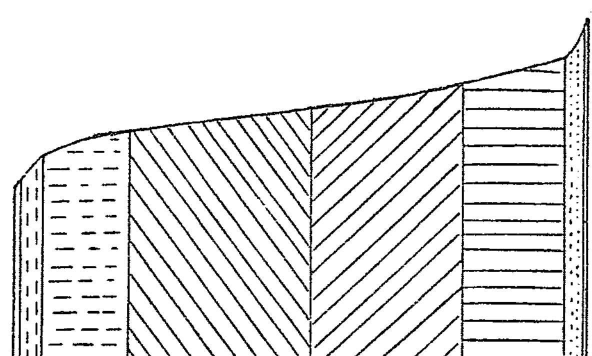
由前各統系觀之，則見其中之個體，與平均數相同者較多，相異者較少，與前二章所得之結果適等，惟表明之方法不同耳。例如「例一」，意謂185顆堅果中，其長度在3·2生的米突與3·6生的米突之間者有7顆，其長度在3·6生的米突與4·0生的米突之間者，有28顆。如是排置之，至所有之果殼，分配於此團體或彼團體而止。換言之，堅果分配爲各種之團體，譬如「隨意」At random 採集185顆堅果，按長短之秩序，立爲等距離之各階級，而分成爲各團體，以各個體分配之，而觀每一團體之受分配者有幾次，如是之分配，統計學家名之曰「次數之分配」。Frequency distributions

以「例一」畫之成圖，則有如第十四圖之形狀。前二章之圖，表明個體，故以直線表明之，此則表明團體，故以方塊 Blocks 表明之。方塊之高度，與團體中所有個體之數成比例。讀者以此圖與下圖比較，則見二圖中所畫黑線之形狀等，則其所處之地位等。例如第十五圖，在4·4生的米突（平均數）與4·8生的米突之間，有四十九個體，而第十四圖中同等黑線之方塊，則與此數成比例。故吾人得曰第十四圖方塊之高度，與第十五圖同黑線之寬度相等。

第十四圖



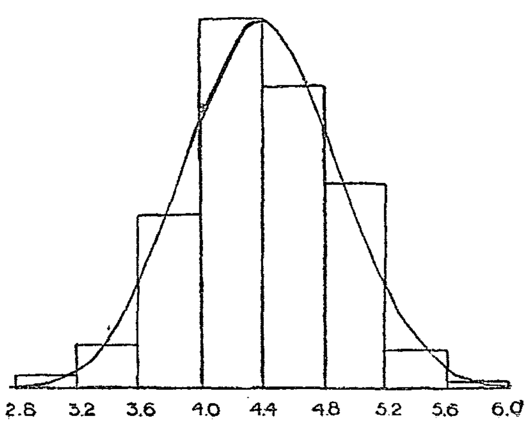
第十五圖



前二章中聯接各直線之上端，可畫一曲線，以觀察所試驗事物變化之概狀若何。同此理也，經過各方塊，可畫一「鐘形曲線」Bell-shaped curve，以觀察所試驗事物變化之概狀若何。

如欲作「鐘形曲線」，請再討論第十四圖。第十四圖，有上下升降之階級 Steps，而成為凹凸不平之「多邊形」Polygon。遠平均數之團體，所包含之個體，較近平均數之團體更少。故苟將所有之任何團體二分之，則近平均數之部分，所包含之個體，較遠平均數之部分必更多。例如「例一」中堅果之長度，在4.0生的米突與4.4生的米突團體之中者，有五十九顆。苟二分之，即4.0至4.2生的米

第十六圖



突爲一部分，4·2至4·4生的米突，又爲一部分，則見前者包含24個體，後者包含35個體。苟將他團體亦二分之，亦呈同一之現象。如是將所有之團體二分之，則「階級」之數愈多，而範圍愈小。分後苟再分之，則階級更愈小。若斯進行不止，則初時凹凸不平之多邊形，終必成爲「和緩之斜面」Smooth slopes 而後已。然非無困難者在。設以「例一」之185顆堅果爲試驗，而二分其所有之團體，惟被分之團體，其中有個體甚少，分後仍不足表明統計所趨之傾向者，此無他，所試驗之物品，爲數不足故也。故欲得完全之「和緩曲線」Perfectly smooth curve 所試驗個體之數，非爲無限之大不可。然無限之大，僅爲理想。苟其大足爲試驗之用，則所成之「和緩曲線」，適爲統計學所欲得之結果。

「和緩曲線」「階級」(卽多邊形)二者，實爲一物，可以第二表證明之。第一行，爲1,000個體在每一小團體所佔之數。第二行，將前行之每一對數併合爲一數。第三行，又將第二行之每一對數，併合爲一數。此爲自左至右，由分而合之排置。苟自右至左，由合而分，亦無不可。惟無論如何，其總數仍爲1,000。

第　二　表

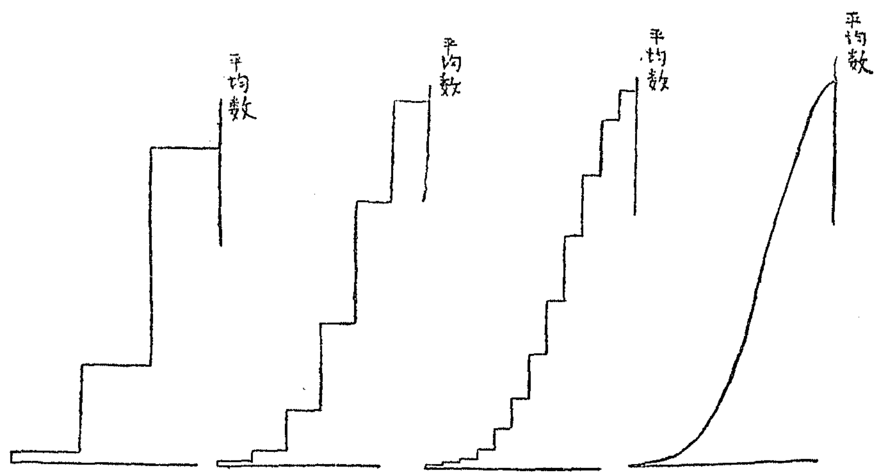
個體之數	併合前行每對之數	併合前行每對之數
1	3	12
2		
3	9	
6		
12	33	117
21		
34	84	
50		
69	156	371
87		
103	215	
112		
平均數	平均數	平均數
112	215	371
103		
87	156	
69		
50	84	117
34		
21	33	
12		
6	9	12
3		
2	3	
1		
1,000	1,000	1,000

第十七圖，表明階級逐漸變爲「和緩曲線」之狀況。惟欲省地位，僅顯半形而已。

讀者當知「階級」屬於人爲。吾人採集統計時，以寸或年歲或斤，擅斷物品之限度，故階級

隨之而起。但寸或年歲或斤之區別，旣全爲擅斷，宜代之與人爲無關係之物。「和緩曲線」無圖

第十七圖



表明「階級」逐漸變爲「和緩曲線」之狀況

體之分，故較凹凸不平之「多邊形」更自然而更普通。

「和緩曲線」以由「次數分配」之結果而得，故又曰「次數曲線」Frequency curve。次數曲線應用之處，請舉一例以明之。試回論第十六圖，設吾人之所知者，僅爲方塊所表明已分成各團體所包含個體之數，然有時吾人欲知在一新團體或如在 4·2 至 4·6 生的米突之間者，所包含個體之數爲幾何，亦意中事。譬如調查戶口之報告，往往以相差十歲爲區別之標準，如三十至四十，四十至五十等是已。但有時吾人欲知三十二歲至四十九歲者有幾人，亦意中事。設在此例，「和緩曲線」已畫成，即自底線 4·2 至 4·6 生的米突之處，畫二垂線至曲線，度量此段曲線之度數，即可知新團體所包含個體之數。但讀者或有疑在此例，曷不自原數計算之，殊不知戶口之統計手續甚繁，欲自原數計算之，勢有所不能。

「次數曲線」另一應用之處，請又言之。本章開端所舉之例，標本之個體皆甚少，今復特舉爲例，而觀察其結果當如何。設「隨意」採取十二個貝殼排置之，使其所分之團體，與本章「例一」相同，則見其分配之情形有如下：

圓體	2·8—	3·2—	3·6—	4·0—	4·4—	4·8—	5·2—	5·6—
十二個標本	0	1	2	1	4	3	1	0
〔例一〕	2	7	28	59	49	33	6	1

觀二統系，頗相類似。但「例一」已非爲完全整齊之統系，而十二個貝殼之標本則更甚。統計學之調查，貴能得貝殼各種之長度。惟所採集之貝殼愈多，則所成之統系愈整齊，終必至絕對的和緩而後已。故苟有方法，自十二個標本或自「例一」之一百個標本，能畫成眞確之「和緩曲線」則以標本少而所起不整齊之形狀，由之而免。且其所得之結果，與度量貝殼全部所得之結果亦較近。惟須注意者，自採集之材料，由手畫成之「和緩曲線」，易與當得之結果，相差甚距。然吾人如何知其可應用之於統計乎？曰，有攷證之二方法。一，由觀察偶然之遭際，如擲錢或擲骰子所當得之結果，而可知「和緩曲線」所當得之形狀。二，由觀察各種之統計所已得之結果，而可知「和緩曲線」所當得之形狀。但此不過略說「和緩曲線」，可以偶然之遭際與各種統計之成例，爲攷證之材料。若夫其詳，則非本書之篇幅所能及。

上所述和緩曲線之大旨，可概括之如下：

一，採集統計之材料，以寸或斤或年歲擅斷物品之限度，凹凸不平之階級，因之而起。「和緩曲線」則足以消除此人為之區別。

二，吾人所試驗之物品，往往苦為數不大，而不能得統計應得之結果。苟能畫成眞確之和緩曲線，則所得之結果，較近於眞相。

第四章 密集數 Mode 標準差數 Standard Deviation 變化之指數 Cofficient of Variation

前言譚臬克力夫在1907年打棍球之點數，有時或多於30，或少於30，平均為30·2。茲復提出二問題而討論之。

一，譚臬克力夫在1907年，每次與人賽球時，其大概所得之點數 Most likely score，當為何數。

二，譚臬克力夫之打棍球，是否前後「一致」Consistent，無大差異乎？

一，請先討論第一問題。吾人初思之，以爲平均數，當爲每次大概所得之點數，然事有未必然者。試設想一不善於打棍球之人，而先後得有以下之點數。

5, 6, 9, 8, 5, 0, 1, 3, 12, 0, 2, 0, 5, 0, 11, 8, 0, 10, 0, 1, 2, 6, 1, 0, 3.

如是打棍球25次，共得98點數，每次平均數爲3·92。苟將其先後所得之點數，按「次數分配」之方法排置之，則有如下之結果。

點數	0	1	2	3	5	6	8	9	10	11	12
次數	7	3	2	2	3	2	2	1	1	1	1

由斯表觀察之，此打棍球之人，從未得4點，二次得3點，七次並無點數，三次僅有1點。若以平均數而言之，較平均數少者十四次，多者不過十一次。若是乎，於彼打棍球之時，欲預料其每次所得之點數，則汝將曰0。何也？以其成例言之，二十五次之中，無點數者，有七次之多也。苟汝曰4，則汝之言中者無一次矣。此爲平均數未必爲「密集數」Mode之明證。（案，密集數之意義云何？

與平均數中位數之比較如何?可作一例以明之:設有一羣人於此,其年齡五十歲者二人,四十歲者四人,三十歲者五人,二十五歲者十人,二十歲者八人,十歲者六人。其「密集數」爲二十五歲,以如此分配之人數中,十人較爲稠密,故以「密集數」名之;其平均數以八百八十歲之總數,爲三十五人所除,爲二十五歲一月有零;其中位數以第十八位之數爲數,爲二十五歲。讀者由此可知三者之區別,「密集數」以比較之關係而起,平均數以數目之關係而起,中位數以位置之關係而起;三者之數,有時有相等者,有不相等者,而以相等者爲多,即在此例,亦幾相等。故「密集數」,平均數,中位數,皆爲度量個體之有用準個。) 惟第十三位之中位數爲3,較「密集數」略近於平均數。今試歸入譚臬克力夫之本題,以好手若彼,所得之小點數尙不少。讀者觀下所列之點數,以彼在1905年1906年1907年三年中之成績爲根據者,可自知之。

點數	0—9	10—19	20—29	30—39	40—49	50—59	60—69
次數	39	25	19	15	7	11	5
點數	70—79	80—89	90—99	100—109	110—119	120—129	130—139

次數	4	3	3	1	2	1	2

當彼打棍球之時，人苟問汝以彼所當得之點數，則汝將曰當屬於第一類。何也？以彼之平均點數，雖在20以上，而其較多之密集點數，則爲0－9也。

欲增助注意於棍球平均數者之興味，今舉數著名好手，在1905年1906年1907年三時令中所得之點數，以先後爲次，列表如下：

海浮 (Hayward)

1905　43，27，13，9，10，34，59，53，13，116，11，4，3，31，22，35，3，24，21，128，168，52，122，17，148，91，88，14，203，13，64，177，88，76，106，112，2，25，48，64，26，24，216，81，58，14，33，32，35，53，10，9，197，12，28，28，0，44，2.

1906　9，37，24，24，58，10，22，255，13，73，1，5，47，3，110，80，0，16，8，21，20，3，31，70，55，26，60，68，32，143，80，5，19，1，115.

82, 70, 52, 6, 85, 6, 33, 31, 6, 59, 2, 66, 71, 63, 44, 8, 40, 122, 29, 38, 76, 32, 17.

1907 39, 82, 25, 5, 219, 135, 0, 46, 0, 68, 10, 31, 194, 244, 143, 125, 54, 69, 6, 15, 100, 144, 34, 144, 208, 54, 34, 76, 10, 16, 10, 109, 51, 18, 14, 76, 168, 17, 3, 94, 45, 51, 8, 9, 24, 60, 110, 141, 22, 28, 62, 51, 17.

戚紹 (Jessop)

1905 0, 0, 4, 20, 61, 26, 8, 87, 10, 29, 10, 52, 5, 27, 40, 12, 2, 11, 34, 206, 48, 43, 0, 1, 1, 3, 7, 27, 77, 17, 43, 42, 24, 20, 23, 10, 68, 170, 58, 47, 77.

1906 28, 1, 60, 38, 14, 48, 0, 0, 1, 25, 1, 17, 20, 11, 14, 89, 234, 5, 23, 0, 1, 52, 4, 2, 1, 4, 19, 15, 54, 21, 65, 57, 60, 8, 10, 15, 11, 1, 45, 0, 74, 8, 25, 0, 43, 0, 1.

1907　63, 55, 5, 17, 16, 42, 8, 8, 87, 27, 22, 34, 4, 48, 5, 80, 66, 40, 0, 15, 6, 12, 111, 0, 4, 19, 53, 4, 6, 0, 48, 15, 75, 26, 26, 28, 1, 10, 34, 2, 14, 9, 54.

譚泉克力夫 (Tunnicliffe)

1905　15, 85, 12, 77, 9, 18, 12, 9, 16, 4, 24, 8, 119, 2, 18, 6, 24, 20, 73, 50, 15, 94, 31, 0, 0, 128, 10, 135, 0, 139, 15, 40, 55, 10, 64, 26, 18, 20, 32, 34, 0, 0, 90, 13, 80, 18, 16, 28.

1906　3, 71, 12, 0, 9, 43, 0, 52, 2, 102, 0, 117, 24, 6, 15, 1, 71, 3, 5, 24, 0, 16, 4, 31, 22, 3, 50, 36, 58, 37, 17, 9, 6, 13, 10, 11, 33, 59, 20, 57, 25, 14, 16, 29.

1907　65, 5, 2, 98, 4, 26, 9, 63, 29, 30, 33, 40, 1, 21, 45, 33, 20, 50, 3, 0, 0, 42, 35, 47, 1, 12, 60, 53, 34, 15, 43, 57, 1, 63, 1, 33, 23, 3, 4, 38, 28, 30, 53, 21, 85.

華鈕 (Warner)

1905 27, 76, 14,13, 47, 45, 39, 7, 15, 34, 38, 106, 107, 75, 8, 3, 4, 4, 4, 47, 13, 209, 66, 14, 0, 78, 10, 23, 0, 0, 44, 16.

1906 0, 10, 204, 0, 0, 60, 0, 85, 0, 49, 86, 3, 50, 5, 10, 7, 70, 43, 9, 59, 97, 0, 166, 152, 13, 22, 34, 82, 14, 14, 35, 56, 48, 23, 31.

1907 61, 42, 137, 10, 41, 72, 0, 22, 53, 66, 4, 73, 28, 17, 16, 122, 48, 5, 87, 26, 9, 66, 7, 9, 17, 44, 60, 12, 66, 2, 77.

是數例者，所顯之特質，大概相同。故得斷言之曰，平均點數，不足爲推測密集點數之標準。

棍球之點數，既如此矣。若夫貝殼或堅果之長度或寬度，其「密集數」爲何，此問題也，與棍球點數之問題無以異。但在此之「密集數，」則與平均數較近，其故安在乎。夫以貝殼或堅果大多

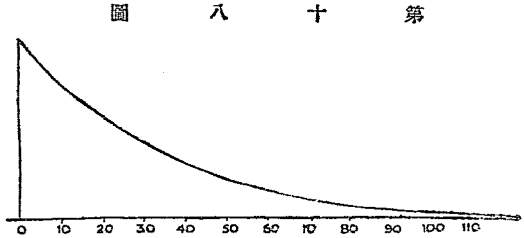
數之軀材，既不易測定，以何法而得度量之乎。曰，二者之疑問，皆可以前章所論之曲線，一一解釋之。

當吾人爲「次數分配」之時，排置統計，分爲各團體。以大小幾相等者，集合之爲同一團體。苟有其中之一團體，較之他團體爲最多，則在曲線中爲最高而卽爲「密集數」所在之部分。試比較第十八第十九二圖之曲線，則見第十九圖爲勻稱之曲線，第十八圖則不然。

夫在勻稱之曲線中，平均數與中位數相等，而曲線在平均數之兩旁，又互相等，此爲吾人在第二章中所已知者。然在勻稱曲線中之「密集數」亦與平均數相等，讀者當不難推想而知之。

第十八圖之曲線，「密集數」在平均數與中位數之左旁。苟採取統系不同之各種統計，畫成許多之曲線，而計算其平均數與「密集數」之差數，及平均數與中位數之差數，則見前者之差數三倍於後

第十八圖



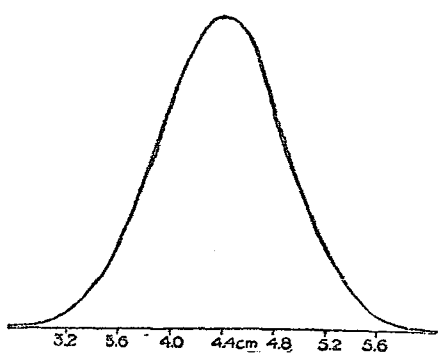
爲棍球點數之曲線表明每一點數所佔之次數密集點數爲○點

者之差數，而中位數在平均數與「密集數」之間。此則為一般之情形，但極端之例，如棍球之點數等則不然，非用更繁雜之方法，不能測定之。然即在如此極端之例，苟欲略知「密集數」所在之位置，則亦不難，由所採集之統計，畫一曲線即得。

二，茲請討論第二問題，即以何方法，而能知上述諸打球名手，是否「一致」是已。

欲解決此問題，則必先知「一致」之意義云何，如何而後能得其價值，而後可。以海浮與戚紹比較，汝見海浮之點數，往往較大，則必以海浮為一致。然不幸而所得之點數常小，亦得謂之一致乎？故僅舉點數之大小，不足為一致與否之證據。設有人於此，永得25點數，而毫無變更，眞可謂絕對的一致。若通常為25，惟有時較多，有時較少，雖不得謂絕對

第十九圖



為堅果之曲線，表明堅果每一長度所估之次數，其最多數之長度，為4.4生的米突

的一致，亦可謂一致矣。換言之，一致之程度，適與差數 Deviation 之多少，成反比例，故斷定一致與否，必以差數之如何為標準。然此猶不足以包括一切也。設甲與乙比較，甲所得之點數，為24，25，26，而乙所得之點數，為34，35，36。僅以差數而言，甲點數之差數為一，乙點數之差數亦為一。但乙之平均點數35，遠大於甲之平均點數25，較甲易失其平均點數，而竟能維持舊狀，故以乙更為一致，於理甚當。設丙與丁比較，丙所得之點數為4，6，8，而丁所得之點數為8，12，16。其差數對平均數而言，則前者之$\frac{4}{6}$，等於後者之$\frac{8}{12}$。平均數之差數，同為$\frac{2}{3}$，故以乙丙一致之程度相等，於理甚當。由此觀之，欲得「一致」之正當價值，不特當論及差數，並當論及差數與平均數之關係，而以平均數為度量差數之標準。

茲先論差數，而後言計算「一致」之方法。讀者當記憶第二章中，曾言由中位數可略知物品之大小，由四分位數可略知大小變化之傾向，是四分位數者，亦為一種差數之表明。然則欲知差數，何不用中位數與四分位數以推測之乎？但有缺點者在。蓋惟勻稱之統計，四分位數之距離中位數，前後相等，否則未有相等者。設有二打棍球之人於此，一則中位點數為35，前四分位點數為

10，後四分位點數爲80。一則中位點數爲37，前四分位點數爲15，後四分位點數爲90。由此四分位數而欲定二者一致之程度，孰高孰下，甚非易易。此四分位數與中位數之距離，前後有不等之困難，不足爲定差數之標準者一。四分位數，對中位數而言，非對平均數而言。但平均數當爲度量差數之標準，既如上之所言。而中位數與平均數，有時相差甚大，此四分位數不足爲定差數之標準者二。顧由此二缺點，而遽謂四分位數爲無用則不可。蓋四分位數本所以表明變化之趨勢，用之以推測差數，往往可省繁雜之算法，而速得價值之大概，功亦不淺。惟四分位數之所表明者，爲統計全體之約略情形，而於個體，並不計及，未免太嫌普通。故欲免其缺點，於是統計學家採用「標準差數」Standard deviation「標準差數」之計算，以平均數爲標準，而計及每一個體之差數。且以每一個體之差數平方之，而消去其負號，使計算便易，尤爲「標準差數」特長之點。

如何而得標準差數，試以華鈕在1907年所得之點數爲例而計算之，有如第三表之結果。第一行，爲華鈕在當年所得之點數，其平均數爲41.9。第二行，爲較平均數或多或少之差數，較多者以正號表明之，較少者以負號表明之。負正號部分之總數，與負負號部分之總數宜相等。第三行，

第　三　表

點　數	平均數減點數之差數	差數平方
61	+19·1	364·81
42	+ 0·1	0·01
137	+95·1	9044·01
10	−31·9	1017·61
41	− 0·9	0·81
72	+30·1	906·01
0	−41·9	1755·61
22	−19·9	396·01
53	+11·1	123·21
66	+24·1	580·81
4	−37·9	1436·41
73	+31·1	967·21
28	−13·9	193·21
17	−24·9	620·01
16	−25·9	670·81
122	+80·1	6416·01
48	+ 6·1	37·21
5	−36·9	1361·61
87	+45·1	2034·01
26	−15·9	252·81
9	−32·9	1082·41
66	+24·1	580·81
7	−34·9	1218·01
9	−32·9	1082·41
17	−24·9	620·01
44	+ 2·1	4·41
60	+18·1	327·61
12	−29·9	894·01
66	+24·1	580·81
2	−39·9	1592·01
77	+35·1	1232·01
		37392·71

$\frac{37392 \cdot 71}{31}=1,206=$差數平方之平均數

$\sqrt{1,206}=34 \cdot 73=$標準差數

第四表

	平均數	標準差數	變化之指數
海浮			
1905	54·9	55·0	100
1906	44·5	43·9	99
1907	66·4	62·0	93
1905–1907	54·9	54·7	99
戚紹			
1905	35·4	42·1	120
1906	26·1	38·5	148
1907	27·9	27·0	97
1905–1907	29·6	36·7	124
譚臬克力夫			
1905	35·8	38·2	106
1906	25·6	27·5	108
1907	30·2	23·9	79
1905–1907	30·2	31·0	103
華鈕			
1905	37·1	43·1	116
1906	43·9	49·3	112
1907	41·9	34·7	83
1905–1907	41·0	43·2	105

為第二行之各數平方而成。若約略之「標準差數」，將前後四分位數與中位數之距離相加，乘之以0·7414即得。以華鈕之棍球點數言之，當為38(眞正之差數為35)但在此例，約略之「標準差數」(苟統計之材料勻稱，較為可恃) 往往不能得極良之結果。

茲將前述四棍球好手三年中打球每年之結果，與三年平均之結果，匯為一表，以供比較。表中變化之指數，Coefficient of variation 所以表明「一致」之程度。一致之程度愈高，則標準差數與平均數之比例愈小。(指數之變化＝100×標準差數÷平均數) 1907年時，氣候諒必乾燥，故此四人者，皆比往年更一致，而譚臬克力夫與華鈕則尤甚。但以三年之平均數而言，海浮不特所得之點數為最高，而一致之程度，亦為最高。

今請舍棍球之點數，而以同一之統計數，解釋堅果貝殼之長寬度。入手初步之方法，可不必重述，僅顯明結果如第五表足已。

由第五表觀之，貝殼堅果之變化，遠少於棍球之點數。讀者苟回思第十八圖第十九圖之情形，即知堅果之個體，與其平均數相差之比例小，棍球點數與其平均數相差之比例大，為不同之

原因。以堅果與貝殼比較，則堅果之變化又更少。

問題之有待於變化之比較而始易解決者讀者不難推想而得。例如動物學家研究各處地方之頭殼，而欲區別此類與彼類之頭殼為同種與否，藉助於指數之比較者不少。

第五章　雙方之關係 Correlation

前者就事物之一方而言，例如堅果僅言其長度或寬度，貝殼亦僅言其長度或寬度是已。若夫長寬「雙方之關係」Correlation 若何，則並未言及。惟「雙方之關係」常為統計學上更重要之問題，本章之所討論者此也。

何謂「雙方之關係，」請先解釋之。設由一堆之堅果中，任取其一，視之則長，吾人能因其長而斷之亦為寬乎？果爾則長寬二者不能謂之無關係。不然，堅果之寬度，不能由其長度

第　五　表

	平均數	標準差數	變化之指數
100個貝殼			
長度（俱以生的米突為單位）	3·10	0·77	25
寬度	1·55	0·34	22
185顆堅果			
長度	4·40	0·54	12
寬度	2·75	0·27	10

推測而知，則二者不能謂之有關係。

至於貝殼，亦可作同一之疑問。然苟堅果貝殼二者之長寬度，吾人皆知其有關係，則當更進一層而作疑問。卽堅果長寬度之關係，與貝殼長寬度之關係，何者更形密切是已。此之所欲爲者，欲於各種之關係，作一比較。有此比較，可按關係深淺之秩序而排列之，猶之吾人於前所作之試驗，按物品長寬度之秩序而排列之也。雖然，欲知長寬度，必先定一尺度，爲度量之標準而後可。欲知關係之程度，亦必先定一種之尺度，爲比較之標準而後可。

試度量堅果，苟其寬度常爲長度之半，而無例外，則堅果之長寬度，可謂有絕對的關係。又或堅果之寬度，常爲其長度之半，而無例外，亦可謂有絕對的關係。換言之，苟二者之關係，常有一定，則其關係不能不謂之絕對的。統計學家定關係之尺度時，以絕對的關係爲準個。所謂準個者，卽「雙方關係之指數」Coefficient of correlation 爲一是已。

假如堅果之長度，約大於寬度二倍，但有時稍多，有時稍少，則關係之程度，幾近於絕對的，而非絕對的也。如此例之價値，較準個（卽一）略小，而爲小數。其數爲何？則以特別之情形而定，祇能

以算術推得之。

前之假設，一則爲有絕對的關係，一則有關係而非絕對的，玆請假設無關係之例而考察之。設一堆堅果，有長短不同之各種大小，已知其平均之寬度爲B生的米突，今又欲知其長度與寬度有無關係。試先採集其最長者而度量之，其寬度爲B生的米突。繼採集其中長者而度量之，其寬度亦爲B生的米突。後採集其最短者而度量之，其寬度又爲B生的米突。如是不能因其長度而推算其寬度，以堅果之長度見告，直不啻曰是爲堅果而已。若此例者，長寬度並無關係，其雙方關係之指數，謂之曰零。

綜上觀之，關係之尺度，可自0至1以表明之，但此尙不足盡所有之例。上所舉之例，言堅果之長者，其寬度或廣或中等耳。然苟事實與之相反，而長者反狹，短者反廣，則其間之關係，不爲長寬而爲長狹。但「狹」不過爲「寬」之反面，統計學家可擴張其計算之範圍，倒退至－1；如是關係之尺度，乃由－1而至＋1。然若以負號爲不便，可以狹度爲標準，而置寬度於不問，指數卽變爲正號矣。

第六表——堅果

		果之寬度											
		1·7—	1·9—	2·1—	2·3—	2·5—	2·7—	2·9—	3·1—	3·3—	3·5—	3·7—	總數
果之長度	3·1—	—	—	2	—	—	1	—	—	—	—	—	3
	3·3—	1	—	—	4	—	—	—	—	—	—	—	5
	3·5—	1	—	—	—	1	1	—	—	—	—	—	3
	3·7—	—	—	2	3	6	6	—	—	—	—	—	17
	3·9—	—	—	1	4	8	7	1	—	—	—	—	21
	4·1—	—	—	—	7	15	4	2	1	—	—	—	29
	4·3—	—	—	—	2	16	6	7	4	—	—	—	35
	4·5—	—	—	—	2	4	7	7	1	2	—	1	24
	4·7—	—	—	—	—	2	8	4	1	—	—	—	15
	4·9—	—	—	—	—	2	5	9	2	1	—	—	19
	5·1—	—	—	—	—	—	3	3	1	1	1	—	9
	5·3—	—	—	—	—	—	—	2	—	—	—	—	2
	5·5—	—	—	—	—	1	—	—	—	1	—	—	2
	5·7—	—	—	—	—	—	—	1	—	—	—	—	1
	總數	2	—	5	22	55	48	36	10	5	1	1	185

推測關係之尺度既定，請轉而討論算法之問題。試先以堅果之長寬度爲例，列爲第六表。表甚明瞭，不待解而自喻。惟其中之數目字，有應說明者，例如第三直行，總計堅果之寬度，在 2·1 與 2·3 生的米突之間者共五顆。其中之二顆，長度在3·1與3·3生的米突之間，又二顆在3·7與3·9生的米突之間，一顆在3·9與4·1生的米突之間。最末之第二橫行，總計長度在5·5與5·7生的米突之間者共二顆。其中之一顆，寬度在2·5與2·7生的米突之間，又一顆在3·3與3·5生的米突之間是。

細觀第六表，以大體言之，果之長者，較果之短者更寬，是長寬有關係可知。苟以堅果之長度爲主，

第七表

堅果之實在長度	前行長度之平均寬度
3·7 cm. 以下	2·31
3·8 cm.	2·58
4·0 cm.	2·63
4·2 cm.	2·63
4·4 cm.	2·77
4·6 cm.	2·90
4·8 cm.	2·85
5·0 cm.	2·95
5·1 cm, 以上	3·04

而計算平均之寬度，則有如第七表。

讀者當知堅果之長度，在3.9與4.1生的米突之間者，作爲4.0生的米突計算。計算平均之寬度亦同。如此計算，雖不免小有謬誤，但所省之煩勞實多。

由第七表以觀，堅果之長度，自3.1增至5.9生的米突，—約增2.6生的米突，（差數實爲2.8，茲以2.6計算者，以兩旁極端之材料甚少故也，計算平均之寬度倣此，）—則寬度亦增0.73生的米突。換言之，寬度增0.73生的米突，同時長度亦增2.6生的米突。

反之以寬度爲主，而計算平均之長度，有如第八表。

若斯平均之長度增1.21生的米突，同時實在之寬度，亦增1.1生的米突。

第八表

堅果之實在寬度	前行寬度之平均長度
2.3 cm. 以下	3.57
2.4 cm.	4.02
2.6 cm.	4.28
2.8 cm.	4.42
3.0 cm.	4.77
3.1 cm. 以上	4.78

由前法所得二比例之效果當如何，試研究之。讀者若記憶前所言「絕對的關係」之理論，則$\frac{0.27}{2.6}$之比例，宜足以度量堅果長寬關係之程度。同此理也，$\frac{1.21}{1.1}$之比例，亦足以度量堅果長寬關係之程度。惟長度與寬度之關係，必等於寬度與長度之關係，今此二比例不等，是不足度量長寬度之關係，而有待於更正也明甚。

以何爲準個而更正此二比例，是亦爲一問題。將以0·2生的米突，同爲長寬度二者之準個乎？抑以0·1生的米突爲一準個，而0·2生的米突又爲一準個乎？曰：物各有其度量之準個，不能强彼以就此。堅果之長寬度，各有其標準差數，爲度量之準個。其長度之標準差數爲0·539，寬度之標準差數爲0·2707，如是長度準個幾爲寬度準個之倍。用之以更正二比例，則其算式爲

實在增加之長度	$2.6 \div 0.54 = 4.82$	
同時增加之平均寬度	$0.73 \div 0.27 = 2.70$	
比　例	$\frac{2.70}{4.82}$	$=0.56$

實在增加之寬度	$1\cdot1\ \div0\cdot27=4\cdot07$	
同時增加之平均長度	$1\cdot21\div0\cdot54=2\cdot24$	
比　例	$\dfrac{2\cdot24}{4\cdot07}$	$=0\cdot55$

若斯二比例幾相等，且顯明堅果長寬度之正當關係，足見算法之無誤。但此算法，不過表示演算之原理，並不以之應用。通常應用之算法，稍形繁雜，其所得堅果長寬度關係之指數，爲0·5742。

另有一簡略之方法，以堅果長寬度之標準差數，作爲長寬度，有如第二十圖。每一長度之平均寬度，以＋號表明之。長度每生的米突之地位，較寬度之每生的米突，適小一半，以其標準差數，較寬度之標準差數，適大一倍故也。圖中斜線之軌道，經過代表平均長度平均寬度二線之交點，又以能接近各×號爲標準。惟須注意者，在二端之×號，不過根據幾個數而算成，故無足輕重。

在如此之圖，而欲推算「雙方關係之指數」以橫垂線至斜線之距離與直垂線至斜線之距

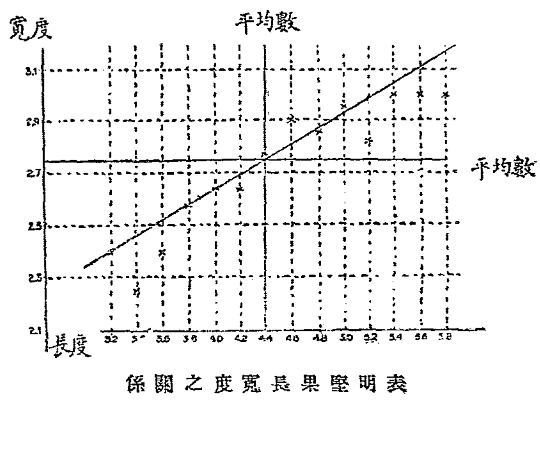
離爲比例卽得。在此例之價値，卽爲上所說之0·5742。

請又舉一例而比較貝殼之長寬度。考察第九表而知其長寬有密切之關係。考察第二十一圖，而知其長寬變方關係之指數爲0·95，幾近於絕對的關係。

讀者觀第二十圖第二十一圖經過各×號之斜線，皆爲直線，恐不免有任何斜線，必爲直線之誤會。然不爲直線而爲曲線者亦有之。但無論何例，直線幾可通用，以相差之數甚微，可置之不問也。

變方關係之指數，用途甚廣，欲了解最新之統計者，不可不知之。

第二十圖



表明堅果長寬度之關係

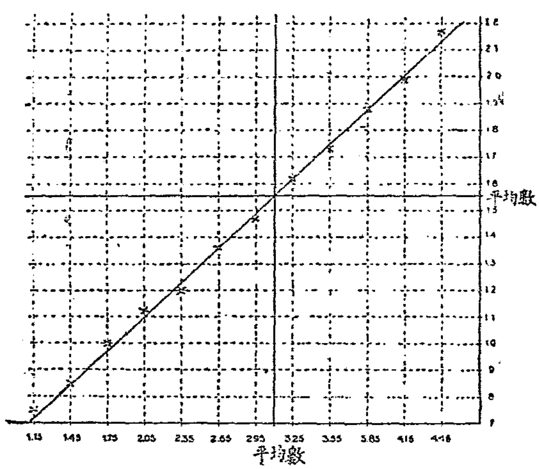
第九表——貝殼

殼之寬度 \ 殼之長度	1·0—	1·3—	1·6—	1·9—	2·2—	2·5—	2·8—	3·1—	3·4—	3·7—	4·0—	4·3—	總數
0·7—	1	1	—	—	—	—	—	—	—	—	—	—	2
0·8—	—	1	—	—	—	—	—	—	—	—	—	—	1
0·9—	—	1	2	—	—	—	—	—	—	—	—	—	3
1·0—	—	—	3	2	1	—	—	—	—	—	—	—	6
1·1—	—	—	—	6	1	1	—	—	—	—	—	—	8
1·2—	—	—	—	—	1	1	1	—	—	—	—	—	3
1·3—	—	—	—	—	1	3	2	1	—	—	—	—	7
1·4—	—	—	—	—	—	4	3	1	—	—	—	—	8
1·5—	—	—	—	—	—	—	4	7	2	—	—	—	13
1·6—	—	—	—	—	—	—	1	7	2	—	—	—	10
1·7—	—	—	—	—	—	—	—	6	8	3	1	—	18
1·8—	—	—	—	—	—	—	—	—	1	4	1	—	6
1·9—	—	—	—	—	—	—	—	—	1	3	1	—	5
2·0—	—	—	—	—	—	—	—	—	—	2	2	—	4
2·1—	—	—	—	—	—	—	—	—	—	—	2	3	5
2·2—	—	—	—	—	—	—	—	—	—	—	—	1	1
總數	1	3	5	8	4	9	11	22	14	12	7	4	100

第十表之立於解釋「雙方關係之指數，」其價值之意義云何不無補助。表中所列人身構造諸關係，(引自麥克湯奈爾博士 Dr. W. R. Macdonell 所著「犯人軀格之研究與犯人之證明」On Criminal Anthropometry and the Identification of Criminals 一論文) 可作爲表明關係程度高下之一尺度觀。其先後之秩序，—

1. 左肘骨與左手中指
2. 全身高度與左肘骨
3. 全身高度與左足
4. 全身高度與左手中指

第二十一圖



表明貝殼長寬度之關係

5。全身高度與面之寬度——為人人意料之所及。蓋左手中指，本為左肘骨之一部分，二者自有密切之關係。身高者之臂短，顯成怪狀，世所鮮有。同是理也，身高者之足，宜不至於短。若有之，較身高而中指短者

第十表

比較之物	雙方關係之指數
上等關係	
貝殼之長寬度	0·95
左肘骨與左手中指	0·85
全身之高度與左肘骨	0·80
全身之高度與左足	0·74
中等關係	
全身之高度與左手中指	0·66
種牛痘之功效與抵抗痘疹之能力	0·64
堅果之長寬度	0·57
頭殼之容量與最大之橫寬度(Naqada男類之頭殼)	0·43
全身之高度與面之寬度	0·35
下等關係	
頭殼之長度與耳之高度	0·29
夫妻之壽命(法國人)	0·22
心臟之重量與人之年齡	0·14
心臟與憂鬱	0·08

自當少見。若夫全身高度，與面之寬度，固有關係。但關係不及上數項之大，可想而知。而表中所列此等關係指數比較之大小，適與事實相合，是以可作爲表明關係程度高下之一尺度觀。惟其中尙缺應有之階級，故復以他種相當之指數補入之。

吾人有所不可不知者，卽關係之眞正原因，非統計所常能表現是已。試以痘症之問題爲例。上級社會，較下級社會，少有痘症之患。欲知其故，試調查二級社會之境況，則見一則多種牛痘而所居之地清潔，一則少種牛痘而所居之地不淸潔。然則二級社會痘症之多少，由於種牛痘與不種牛痘乎？抑由於居地之淸潔與不淸潔乎？或以爲富民所居之地，與外間少往來，故不易有傳染病；貧民所居之地，汚穢稠密，故傳染病盛行。是則種牛痘之預防，不足爲貴，而居地之淸潔，實有莫大之關係。是說也持之有故，言之成理。但據歷次之調查，而種牛痘之功效，彰明昭著，已無容疑。故就痘症之問題而論，上等社會，多種牛痘，而所居之地清潔。下等社會，少種牛痘，而所居之地不淸潔，此皆爲事實。然若根據之以解釋與痘症有無關係，結果或至大不相同。是採集統計之方法雖善，苟解釋之不當，仍不能得良好之結果也。

第六章　或有的差誤 Probable Error

考察同類之事物，所得各標本之平均數，以第一章所得之原理推測之，知其相差不多，但又知其並非完全一樣，然則所謂平均數者不足恃，而吾人之計算，亦不足爲憑乎?，曰，是不然。讀者當曾記憶個體與中位數之差數，在第二章中，吾人知可由四分位數測定之。在第四章中，吾人知可自更良之算法曰「標準差數」者測定之。若標準差數，足以測定堅果或貝殼個體大小變化之情狀，其足以測定同物各標本所得平均數大小變化之情狀也明甚。蓋同整列中個體之標準差數，既可計算而得，則各平均數之標準差數，亦可計算而得。即以計算個體標準差數之法，而應用之以計算平均數之標準差數是已。其法即算得許多之平均數，以排列堅果或貝殼個體之法排列之，而測定其標準差數，則知各平均數變化之差數，當爲若干。若斯計算尋常標準差數，與計算平均數之標準差數，二者之原理同。所不同者，彼則以個體爲主體，此則以平均數爲主體耳。請舉一例以明之，試擲六枚錢二十回，以觀其所有「陽面」之平均數若干，如是者四十五次，其第一次試驗之結果如下。

陽面之數	1	2	3	4	6
所得陽面之次數	2	5	7	5	1

如是第一次陽面之平均數爲2.95，設第二次之平均數爲3.05，第三次又爲3.05，若斯繼續進行，直至吾人得第四十五次之平均數，然後將四十五次全部之平均數，按照第三表之原理排列之，有如第十一表。

第十一表，顯明平均數之標準差數。但讀者當知吾人測定一平均數時，計算不能如此之繁雜，故不得不別求一簡便之方法。

試回觀第一章之各圖而考察之，苟其標準差數愈大，—卽中位數與四分位數之相差愈大，—則聯接各直線上端之線，傾斜愈甚，而知其個體之變化愈大。反之標準差數不大，則聯接各直線上端之線，不甚傾斜，而知有許多相等之個體。若有許多之個體，與中位數相等，而移動中位數四五位於左方或右方，則於中位數之價值，當無甚影響。反之與中位數相等之個體不多，則四五位之移動，於中位數之價值，大有出入。

第十一表

擲六枚錢二十回所得之平均數	次數	平均數(3·0)減第一行各數之差數	差數平方後又乘次數
2·25	1	−0·75	0·56
2·55	1	−0·45	0·20
2·6	2	−0·4	0·32
2·7	2	−0·3	0·18
2·75	4	−0·25	0·25
2·8	3	−0·2	0·12
2·85	3	−0·15	0·07
2·9	3	−0·1	0·03
2·95	2	−0·05	0·00
3·0	4	0·0	0·00
3·05	4	+0·05	0·01
3·1	4	+0·1	0·04
3·2	1	+0·2	0·04
3·25	2	+0·25	0·12
3·3	4	+0·3	0·36
3·4	3	+0·4	0·48
3·55	1	+0·55	0·30
3·65	1	+0·65	0·42
平均數之總數	45		3·50

$$\text{差數平方之平均數}=\frac{3\cdot50}{45}=0\cdot078$$

$$\text{平均數之標準差數}=\sqrt{0\cdot078}=0\cdot28$$

然又有一事，當注意者。假如堅果大小之平均數，自五千顆計算而得，則吾人以爲當可深信。若僅自五顆計算而得，則吾人以爲未必可恃。是平均數之眞確與否，不特與標準差數有關係，如上節之所云云者，且與個體之總數亦有關係。惟所謂關係者，並非謂標準差數，當以個體之總數除之，乃謂標準差數，當以個體總數之平方根除之也。故個體之總數，若多四倍，則平均數之眞確，亦多一倍，個體之總數多九倍，則平均數之眞確，亦多三倍。

苟以此原理，應用之於擲六枚錢二十回之例，則有如下之結果。設以第一次試驗，爲主體之試驗，而以其他之試驗，爲參考之試驗，則見第一次試驗的個體之標準差數爲1·16，而以個體總數(二十)之平方根除之，其所得平均數之標準差數爲·27，與第十一表之得數甚近。

故應用之試驗平均數或其他統計數 Statistical function 之後，往往附之以該數同塈列中個體之標準差數。試以堅果爲例，185顆堅果長度之平均數爲4·40生的米突，個體之標準差數爲0·49生的米突，故平均數之自身，恐不免有0·036生的米突之差（$0·49 \div \sqrt{185} = 0·036$）。通例以0·6745乘前數（即0·036生的米突）爲0·024，而名此得數曰「或有的差誤」。

Probable error

兩平均數相差，往往有之，但相差之數，當不過於「或有的差誤」之三倍。故用「或有的差誤」，可以證明平均數差數之大小如何，始爲適當。讀者須知無論何種之統計數，未能免「或有的差誤。」以「隨意」採集同物各種之標本，往往不能得同一之結果也。然其相差之數，不過於「或有的差誤」之三倍，可斷言者。不然，比較二人種之頭殼，苟無此以爲證明之方法，則相差之數太大，或所取之標本爲數太小，皆無由而知之。

「或有的差誤」最有用之處，卽足以證明雙方關係指數之爲適當與否是已。例如測定二物之關係指數爲0·1時，關係之程度如此之淺，不免起二種之疑問。(一)或以爲眞有關係，惟所取之標本爲數甚小，故其所顯明之關係不大。(二)或以爲實無關係，惟所取之標本爲數甚小，故作有關係之表示。二者孰是孰非，可以「或有的差誤」證明之。「雙方關係指數」之「或有的差誤」其算式爲$0·6745\times\frac{1-r^2}{\sqrt{n}}$。算式中 r 爲指數，n 爲所取標本個體之數。（苟標本個體之數小，此公

式卽不適用。如爲數小於二十，則不能得適當之結果，在二十與三十之間，所得之結果，仍不能視爲確實）以此算式應用之於此例，如所取標本個體之數爲36，則「或有的差誤」爲0·1；較·01之指數反大，斷無是理。若個體之數爲100，則「或有的差誤」爲0·06，雖較指數小，然相差不甚大，故0·1之指數，究足以表明二物之有關係與否，尙屬疑問也。

前章所舉堅果貝殼「雙方關係指數」之「或有的差誤」列表如下。

比較之物	指數	或有的差誤
185個堅果之長寬度	0·57	0·03
100個貝殼之長寬度	0·95	0·06

由此表觀之，堅果與貝殼之長寬度，皆有關係，顯然可見。

讀者當知本章之理論，全以討論之標本「隨意」採集爲前提。卽全書之理論，亦以討論之標本，「隨意」採集爲前提。（案 Selection at random 譯爲「隨意」採集，「隨意」二字，爲統計學之

專門名詞，取其爲「故意」之反對「故意」有成見「隨意」無成見也，如採集極多之葉片，不能一一考察之者專取其最大者或最小者以爲標本，不啻胸有成見，所得者限於一部分，不足以代表全體，是謂「故意」之選擇；若將其全部攪亂之使之大小散處無定所，然後順手取之，則所得者，不偏不頗，必足以代表全體，是謂「隨意」之選擇；若以他方法採集之，不免徒勞而無功。設欲知一人種之平均高度，而集合長人以爲考察之標本，不啻南轅而北轍。設欲知堅果之大小而自高價值所得之果，則其長寬度之價值亦必高；自底價值所得之果，則其長寬度之價值亦必底。二者或趨於極端，其平均數之差，竟數倍於「或有的差誤」。然既非出自異種，又非來自異地，而二者大相懸殊，此無他，採集不適當之過也。

　　請又舉一例：設欲調查境遇之窮困與個人殘疾之關係，而至倫敦病民之區域，徒計算其殘疾之人數，而不及其他；又或至富民之區域，徒計算其殘疾之人數，而不及其他。其所得之效果，直等於無用。何以故？以無比較故。設殘疾之人，前者之數爲20,000，而後者之數爲5,000，苟不知其每一區域之人口幾何，則仍無所證明，故必知貧民區域之人口爲 2,000,000，富民區域之人

口爲500,000,而後知二級社會,表示同一之比例。此爲簡單之例,證明標準之採集,非「隨意」不可。不但殘疾之人,爲吾人所當知,卽非殘疾之人,亦爲吾人所當知。蓋有比較,斯有區別,有區別,始有顯然之結果。又欲有所言者,與其取50,000故意選擇之個體爲標準,遠不若取5,000或6,000隨意採集之個體爲標準。何也?一則爲數雖多,而無所表示,一則爲數雖少,而有所表示也。吾人之對於此點,斷斷不已者,以見夫世之人,有採集特別之標本,積歲累月,而不值爲統計學家之材料,此蓋爲採集統計最普通之缺點,亦爲最危險之缺點。讀者苟欲爲統計之調查,或欲評論統計之著作,無論統計之題目,爲人爲獸爲貝殼爲堅果或爲棍球之點數,當注意之第一着,其材料必須採集自全部,而以「隨意」得之,其審查之態度,必出之以無黨無偏而後可。

Primer of Statistics

Commercial Press, Limited

中華民國十二年二月初版

(統計學原理一册)
(每册定價大洋貳角)
(外埠酌加運費匯費)

著者 W. Palin & Ethel M. Elderton

譯者 嵊縣趙文銳

發行者 商務印書館

印刷所 商務印書館 上海北河南路北首寶山路

總發行所 商務印書館 上海棋盤街中市

分售處 商務印書館分館 北京天津保定奉天吉林龍江濟南太原開封鄭州西安南京杭州蘭谿安慶蕪湖南昌漢口長沙常德衡州成都重慶瀘縣福州廣州潮州香港梧州雲南貴陽張家口新嘉坡

六一〇四