MAN IN INDIA.

Vol. VIII. APRIL & SEPT. 1928. Nos. 2 & 3.

I.—A STATISTICAL STUDY OF THE CHINESE HEAD

BY

P. C. MAHALANOBIS

- 1. The material for the present study consists of two series of measurements on living men taken by S. M. Shirokogoroff (Anthropologist, Russian Academy of Science) in 1908-12 and 1923-24 respectively, and published by the North China Branch of the Royal Asiatic Society in the form of two reports:—
 - (i) "Anthropology of Northern China, etc." Extra Vol. II, Shanghai, 1923.
 - (ii) "Anthropology of Eastern China, etc." Extra Vol. IV, Shanghai, 1925.

Shirokogoroff classified his material according to the province of origin (birth place) of the person examined, and in the earlier portion of my comparisons I have retained his classification. The following list gives the names of the different provincial groups. The number of individuals available in each case is given within brackets.

The so-called "non-selected" data in every case consist of persons measured in hospitals and asylums, while the "criminal" data belong to inmates of Shanghai and Hong-Kong Municipal gaols. It will be noticed that although Shirokogoroff uses the name "non-selected" such data do not really represent a random sample of the general population but are selected with reference to disease and disorders of the mind.

- (i) Manchuria (96).
- (ii) Chihli (114).
- (iii) Shantung (185).
- (iv) Kiangsu "non-selected" (102).
- (v) Kiangsu "criminal" (113).
- (vi) Chekiang "non-selected" (62).
- (vii) Chekiang "criminal" (44).
- (viii) Anhwei "non-selected" (44).
 - (xi) Kwangtung "non-selected" (110).
 - (x) Kwangtung "criminal" (220).

In addition to the above Chinese samples I have also taken (xi) Koreans (141), consisting of men from North Korea and the maritime provinces of Siberia, and (xii) Manchus (81), from the Aigun district of Heilung-Kiang.

2. In the present paper I have considered only the measurements on the head (flesh), altogether 16 in number. Shirokogoroff gives the mean values for each province, but does not usually give individual measurements or standard deviations. I have therefore been obliged to use standard deviations calculated from a long series of measurements of 550 Korean men published by T. Kubo.¹

Mean values of groups discussed in the present paper are given in Table I(A).

^{(1) &}quot;Beiträge zur physischen Anthropologie der Koreaner" (Mitt. med. Fakult. Kais.-Univ. Tokio, Bd. xii, 1913).

TABLE I (A) Mean values for Chinese Samples (Shirokogoroff).

187.9 " 186.4 " 186.4 34.3 " 35.3 " 34.7	:::	184 104.8 " 104.6 " 185 109.3 " 187.5 " " 184 191.6 " 185.3 " " 34.3 " 35.6 " "
, 135·3 , 133·4 , 134·0		35-3 " 94-6 185 134-1 " 134-1

TABLE I (B).

Mean values of certain samples from Northern Asia and China (Shirokogoroff). The Chinese groups are obtained by pooling together some of the provincial means given by Shirokogoroff.

Character.	Kor	eans.	Mar	nchus.		thern nese.		stern nese.		thern inese.
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Head Length .	141	183·1	81	181.9	297	187.8	365	186:4	33 0	185 2
Head Breadth .	,,	153.7	80	151.3	,,	148:3	,,	152·1	,,	148.0
Morphological Face Length.	,,	117.2	,,	117.7	,,	117.8	,,	115.8	,,	112.5
Bizygomatic Breadth.	,,	143.8	79	140:3	,,	140.5	,,	141.9	,,	137.9
Nasal Height .	,,	40.8	81	45.4	296	41.6	,,	42.7	,,	40.0
Nasal Breadth .	,,	37.3	81	37.9	238	37.2	,,	37.3	,,	37.0
Cephalic Index .	,,	83.7	80	83.2	297	79.1	,,	81.7	,,	79.9
Nasal Index .	,,	92.5	81	83.0	,,	89.9	,,	87.5	,,	93.2
Morphological Facial Index.	140	81.6	79	83.9	295	83.9	,,	81.7	,,	81.6
Minimum Frontal Diameter.	141	106:3	80	105.4	298	104.7	,,	104.9	,,,	103.4
Bigonial Diameter	140	112:3	,,	110.5	299	109.2	,,	107.2	,,	105.3
Physiological Face Length.	141	194.7	79	188.3	296	192.1	,,	186.6	,,	183.6
Internal Ocular Breadth.	142	33 ·9	81	34·1	297	34.3	,,	34.9	,,	34.0
External Ocular Breadth.	"	95·3	,,	93.5	,,	95.6	,,	92.2	,,	88.1
Height of Head .	,,	134.5	,,	132.6	298	134.6	,,	134.3	,,	131.5
Ear Length .	141	63.9	,,	65.0	296	64.1	,,	61.2	,,	59.6

3. We shall first test whether the different provincial groups may be considered to be samples drawn from the same general population, or whether they must be considered to be statistically divergent, i.e. belonging to different populations. I have used Prof. Pearson's C² test for this purpose², and I give the values of C² for the 45 different pairs of groups in Table II.

The mean value of C^2 for two groups which belong to the same population is zero with a probable error of $\pm .67449 \sqrt{\frac{2}{p}}$, where p is the total number of characters. In the present case p=16, and therefore the probable error $=\pm 0.24$ approximately in every case.

The test therefore consists in comparing the observed values of C^2 with a theoretical value of 0 ± 0.24 . If the observed value does not differ from zero by more than say 0.72, then the two groups may be considered to be drawn from the same general population; on the other hand if the observed value is greater than 0.72, then the chances are that the two groups belong to different populations.

A glance at Table II will show that all the coefficients are significantly greater than 0.72, with only two exceptions, namely, Chihli and Shantung with a coefficient of 0.65 ± 0.24 ; Kiangsu criminal and Chekiang criminal with a coefficient of $0.54 \pm .024$. We also notice that comparatively high values of C^2 occur several times in Table II, e.g., 53.79, 45.04, 45.78, etc.

We conclude that speaking generally the different provincial groups must be considered to be significantly differentiated from one another.

⁽²⁾ Biometrika, Vol. 18, 1926, pp. 105-117.

TABLE II.—Values of C² for Chinese samples.

				•			•				
				5	Kr	KIANGSU.	Снек	CHEKIANG.	44) u-sej-	Kwaz	Kwantung.
		situdo na ld (88)	(113) CP!PI	(181)	-\$59leston" (201) "be	"lanimi10" (EII)	toelea-non'' (53) ''bo	"[animh10" (44)	on" iewdnA	-toelea-non'' (011) ''1.e	lsalmi10" (022)
Manchuria (96)		:	5.55	11.05	6.46	99.8	6.51	99.6	6.46	43.78	37.48
Chihli (113)		5.55	•	0.65	12.97	11-70	6.19	10.35	9-22	42.04	34.60
Shantung (184)		11.05	0.65	:	16-45	13.11	6.05	6.67	12.40	53.19	38.20
Kiangsu "non-selected" (102) .		97.9	12.97	16.45	:	3.84	1.08	4.99	2.85	33-39	23.67
Kiangsu "criminal" (113)	-	99.8	11.70	13.11	3.84	:	17.1	0.54	1.83	28.37	18.91
Chekiang 'non-selected'' (62) .	•	6.21	6.19	6.05	1.08	1.71	:	2.58	2.41	19.17	10-20
Chekiang "criminal" (44).		99.6	10.35	10.0	4-99	0.24	9.58	:	1.67	98.6	3.68
Anbwei (44)		95-9	9.32	12:40	2.85	1.8:3	7:	1.67	:	20.59	8.22
Kwantung "non-selected" (110)	-	43-78	45.04	53.79	33-30	28.37	19-17	98.6	20.59	:	5.38
Kwantung "criminal" (220)		37-48	34.60	38.29	23-67	18.91	10.50	3.68	8.55	5.38	:
	ı										

Probable error is in every case ± 0.24 approximately.

4. I shall now proceed to discuss the actual magnitudes of the divergence between different groups. I have used for this purpose a certain coefficient of divergence defined by the following equation:—

$$D^{2} = \left[\frac{\mathbf{I}}{p} \mathbf{S} \frac{(M_{p} - M'_{p})^{2}}{\sigma_{p}^{2}}\right] - \left(\frac{n + n'}{n \cdot n'}\right) \dots (\mathbf{I})$$

with a variance given by

$$\Sigma^{2} = \frac{4}{p} \left(\frac{n+n'}{n \cdot n'} \right) \cdot \overline{D}^{2} + \frac{2}{p} \left(\frac{n+n'}{n \cdot n'} \right)^{2} \cdot \cdot \cdot \cdot \cdot (2)$$

where M_p , M'_p are the observed mean values of the pth character in two groups of size n, n'; ρ_p^2 is a reliable value of the variance of the pth character which is kept constant throughout the whole series of comparisons, and \overline{D}^2 is the mean value of D^2 . The summation extends over all characters, the total number of which is given by p.¹

Table III gives \widehat{D}^2 (together with the probable errors) computed in accordance with the above formulæ. I have used the standard deviations of a long Korean series for σ_p in equation (1), and observed values of D^2 for mean values \widehat{D}^2 in equation (2).

I need hardly mention that D^2 measures divergence between two groups, i.e., the greater the value of D^2 the greater is the divergence, while the smaller the value of D^2 the greater is the resemblance between the two groups.

5. The "non-selected" (i.e. hospital and asylum inmates) and the "criminal" groups from the same province all give very low co-efficients. For example, Kiangsu $.071\pm.017$, Chekiang $.094\pm.024$, and Kwangtung $.074\pm.011$. The divergence is statistically significant but small in magnitude in every case.

⁽¹⁾ The theoretical foundations of equations (1) a:.d (2) have been discussed by me in a separate paper which will be published shortly.

TABLE III.—Values of D² for Chinese samples.

	Manchuria (96)	Chihli (113)	Shantung (184)	Kiangsu "non-selec- ed (102)	Kiangsu "criminal" (113)	Chekiang "non-selec- ted (62)	Chekiang "criminal" (44)	Chekiang Anhwei Kwangtung Kwangtung **criminal" **inon-selec. "non-selec ''roriminal" (44) ted" (110) (220)	Kwangtung "non-selec- ted" (110)	Kwangtung "oriminal" (220)
Manchuria (96)	:	107±-017	175±.018	081±.015		·168±·020 ·173±·024 ·334±·037 ·215±·029	-334±-037	215±.029	.935±.046	.563±.031
Chibli (113) .	107±-017	:	.010±-006	.237±.023		·207±·021 ·155±·023	.327±.036	.353±.011	-843±-042	•464±•027
Shantung (184)	175±018	900.∓010.	:	.251±-021	188±-018	131±.019	.281±.031	.350±.011	-746 ± 035	-360±-020
Kiangsu (102). 'non-selected":	-081±-015	$.081\pm.016$ $.237\pm.023$ $.251\pm.021$.251±.021	:	·071±·017	·030±·013	.162±.027	·030±·013 ·162±·027 ·093±·021 ·630±·037	-630±-037	-342±-036
Kiangsu (113). "criminal"	168±.020	-207±-021)20 -207±-021 -188±-018 -071±-017	·071±·017	:	·047±·013	.018±.013	·047±·013 ·018±·013 ·059±·018 ·497±·018	-497±-018	·212±·018
Chekiang (62). 'non-selected"	-173±-024	•155±•023	024 ·155±·023 ·131±·019 ·030±·013 ·047±·013	.030±.013	·047±·013	:	.094±.024	·094±·024 ·093±·025 ·484±·038	.484±.038	-204±-022
Chekiang (44). "criminal"	-334±-037	·327±·036	·281±·031	·162±·027	337 -327上・036 -281上・031 -162土・027 -018土・013 -094土・024	.094±.024	:	.083±.026	.314±.036	.102±.020
Anhwei (44). 'non-selected"	-215±-029	-353±-011	.350±.011	.003±.021	·353±·011 ·350±·011 ·003±·021 ·059±·018 ·093±·025		.083±.026	:	.446±.041	.227±.028
Kwantung (110).	-935±-046	-843±-045	·746±·035	·746±·035 ·630±·037 ·497±·018		.484±.038 .314±.038 .446±.041	.314±.036	.446±.041	:	110-∓-280-
Kwantung (220). "criminal"	.∓893-)31 ·464±·027		.342±.036	・360±・220 ・342±・036 ・212±・018 ・204±・022 ・102±・020 ・227±・028 ・087±・011	204±.02°	.102±.030	·227±·028	110-∓280.	:
			_			_			_	

- 6. Chihli and Shantung are both situated in the north, and naturally enough they have a coefficient of only+.010±.006 which is negligibly small. We have also seen that as judged by Pearson's C² test these two groups may be considered to be drawn from the same population. I have therefore clubbed them together to form a single "Northern Chinese" sample for later comparisons.
- 7. The 5 samples from the eastern provinces of Kiangsu ("non-selected" and "criminal"), Chekiang ("non-selected" and "criminal"), and Anhwei ("non-selected") all with one exception give coefficients less than .100 (values of C^2 less than 5.0), and therefore exhibit comparatively high association or resemblance with one another. The case of Kiangsu "non-selected" and Chekiang "criminal" with a coefficient of + .162 \pm .27 ($C^2=4.99$) is the only exception; but even here the value of D^2 is not significantly greater than + .100.

All these groups may therefore be considered to belong to a slightly generalised "Eastern Chinese" population, and I have accordingly pooled them together to form one such group.

8. Kwangtung (which lies much further south) appears to be definitely divergent from the other northern provinces. Here I have clubbed together the "non-selected" and "criminal" samples (with a small coefficient of + .074±.011) into one single group of "Southern Chinese."

In this connection I may note a curious fact that Kwangtung "criminals" show distinctly greater resemblance than Kwangtung "non-selected" with every other sample. Prof. Karl Pearson to whom I had shown this

result suggested as an explanation that "the "criminals" in this case included a larger proportion of men from other provinces (who for some reason or other wanted to keep dark their real province of origin), just as a large proportion of the criminals of London are foreigners."

9. I shall now compare the Manchus (80), the Koreans (141), and the Chinese of Manchuria (96) with the pooled samples of "Northern" (297), "Eastern" (365), and "Southern" (330) Chinese.

Table IV gives the observed values of Pearson's C^2 . The theoretical value (on the assumption of no divergence) is $0\pm.24$. It will be noticed that all the coefficients are significantly greater than zero, showing that all the groups may be considered significantly divergent.

Table IV.—Value of C^2 for Chinese samples. (Probable error = ± 0.24 .)

	Koreans (141)	Manchus (80)	Chinese of Manchu- ria (96)	Northern Chinese (297)	Eastern Chinese (365)	Southein Chinese (330
Koreans (141)		18.10	3.97	20.68	25:37	75:75
Manchus (80)	18.10		5 89	17.71	13.12	43.58
Chinese of Manchuria	3.97	5.89		10.45	11.64	49.19
(96). Northern Chinese (297)	20.68	17.71	10.45	[32.75	80.64
Eastern Chinese (365).	25.37	13.12	11.64	32.75		49.58
Southern Chinese (330)	77.75	43:58	49.19	80.64	49.58	<u> </u>

10. Table V gives the values of D² together with corresponding values of the probable error.

Koreans show very great resemblance with the Chinese of Manchuria (.070±.013), and moderate

Table V.-Values of D² for different Chinese samples.

			ų E	CHINESE.	£.	
	Koreans (141)	Manchus (80)	Manchuria (96)	North (297)	East (365)	South (330)
Koreans (141)	:	.348±.028	.070±.013	910. ∓ 212.	- - - - - - - - - - - - - - - - - - -	.767±.029
Manchus (80)	-348±-028	:	·136±·020	.280±.023	810. T 661.	.675±·035
Coinese Manchuria (96)	.070±.013	·136±·020	:	·144±·015	·154±·015	·664±·018
Northern Chinese (297)	910.∓.010·	.280±082	.144±.015	:	.200±.012	·517±·019
Eastern Chinese (365)	.284±.018	\$10. T 661.	.154±.015	-200±-012	:	.285±.013
Southern Chinese (330) .	·767±·029	.675±·035	810.∓ 1 99.	610. ± 212.	£10.∓582.	:

but quite appreciable association with the Northern $(.217\pm.016)$ and Eastern $(.284\pm.018)$ Chinese. The difference between the two latter coefficients $(.067\pm.025)$ is not significant, which is not surprising since these two regions are actually contiguous.

Manchus (.348 \pm .028) are substantially differentiated, while the divergence from the Southern Chinese (.767 \pm .029) is still more marked.

- 11. Manchus show fairly close resemblance with the Chinese of Manchuria ($\cdot 136 \pm \cdot 020$). Association with the Northern ($\cdot 280 \pm \cdot 023$), and the Eastern ($\cdot 199 \pm \cdot 018$) Chinese is also quite appreciable, the difference between the two latter coefficients ($\cdot 081 \pm \cdot 029$) being again negligible. Divergence from the Koreans ($\cdot 348 \pm \cdot 028$) is moderate, but is pronounced in the case of the Southern Chinese ($\cdot 675 \pm \cdot 035$).
- 12. The Chinese of Manchuria show great resemblance with Koreans (070 ± 013), suggesting strong intermixture or intense convergence, and very appreciable resemblance with Manchus (136 ± 020), Northern Chinese (144 ± 015) and Eastern Chinese (154 ± 015). The Chinese of Manchuria would thus appear to have been derived largely from the Northern and Eastern provinces. They show marked divergence from the Southern Chinese ($.664\pm.018$).
- 13. Northern Chinese. Resemblance is greatest with the Chinese of Manchuria $(0.144\pm.015)$ and is fairly close with the Eastern Chinese $(.200\pm.012)$, Koreans $(.217\pm.016)$ and Manchus $(.280\pm.023)$. But divergence from the Southern Chinese $(.517\pm.019)$ is again quite marked.
 - 14. The Eastern Chinese occupy an intermediate

region and quite naturally show fairly close resemblance with all the other groups. Association is closest with the Chinese of Manchuria ($\cdot 154 \pm \cdot 015$), and is almost equally great with both the Northern Chinese ($\cdot 200 \pm .012$) and the Manchus ($\cdot 199 \pm \cdot 018$). The resemblance with Southern Chinese ($\cdot 285 \pm \cdot 013$) and the Koreans ($\cdot 284 \pm \cdot 018$) is only a little less in degree, but is still quite pronounced.

- 15. The Southern Chinese are markedly divergent from most of the other groups. They are almost equally differentiated from Koreans ($\cdot 767 \pm \cdot 029$), the Manchus ($\cdot 675 \pm \cdot 035$), and the Chinese of Manchuria ($\cdot 664 \pm \cdot 018$), and only to a slightly lower degree from the Northern Chinese ($\cdot 517 \pm \cdot 019$). They show however an appreciable degree of resemblance with the Eastern Chinese ($\cdot 285 \pm \cdot 013$), which is not surprising as the eastern provinces are situated fairly close to the province of Kwangtung from which the southern group is drawn.
- 16. We thus see that all the Chinese groups from the northern provinces e.g. the Chinese of Manchuria, Northern Chinese and Eastern Chinese are closely associated with one another, and all show fairly close resemblance with both Manchus and Koreans who also come from the north. The Southern Chinese on the other hand are clearly differentiated from practically all the northern groups, with the single exception of the Chinese from the eastern provinces (which are adjacent to Kwangtung) with whom they show fairly close association.

Koreans and Manchus, although both show appreciable resemblances with all the Chinese groups from the north, are distinctly differentiated from each other.

They however resemble each other more closely than either of them resembles the Southern Chinese.

17. If we confine our attention to the Chinese samples and look at Table V as a whole and compare it with a map of China, we perceive a very simple relationship between geographical proximity and physical resemblance: the smaller the distance between any two regions the greater is the resemblance between the inhabitants of those two regions, or the greater the distance between any two regions the greater is the divergence between the inhabitants of those two regions.

The Northern and Southern populations are in fact highly differentiated from each other, the change occurring gradually through the eastern provinces lying in the centre.

18. It will be interesting to compare how the different characters vary from group to group. One way of doing this would be to determine the *inter-class* (or "extra-group" as it may be more conveniently called) standard deviations for the whole family, and compare these inter-class standard deviations with the corresponding intra-class S. D.'s.*

$$\Sigma^2 = \frac{1}{q} S (M - Mq)^2 \dots (3.0)$$

^{*} The inter-class (or extra-group) variance is defined for any particular character by

where M is the average for the whole family, and M_q is the mean for the qth group, and the summation extends over all q groups. It will be seen from the above definition that Σ_p represents the average separation (for the pth character) of each group-mean from the general mean. If we denote by σ_p a reliable *intra*-class standard deviation, then it represents the average separation of an individual from its own group-mean. Σp thus represents the variation within the group. As already mentioned I have used throughout Korean values of intra-class variances.

Table VI gives the average value of the ratio $2 \Sigma^2/\sigma^2$ for each character separately (for all the 12 provincial samples discussed in the present paper). Remembering that the mean value of $(M-M')^2$ is simply $2 \Sigma^2$, we notice that $2 \Sigma^2/\sigma^2$ will give the mean value of $D^2 = (M-M')^2$ for the whole family for any particular character. The quantity $2 \Sigma^2/\sigma^2$ therefore gives the average value of the coefficient of divergence (neglecting the small correcting term for the size of the samples) for any character for the family as a whole.

Table VI.—Inter-class and intra-class standard deviations.

			(Σ^2) Interracial variance.	(σ²) Intraracial variance.	$\frac{\sigma^2}{2 \sum^2}$	$\frac{2 \sum^2}{\sigma^2}$
Head Length			4.823	50:341	5.219	0.191
Head Breadth			5.167	31.066	3.007	0.333
Morphological Len	gth		4.436	34.137	3.849	0.260
Bizygomatic Bread	dth		3.541	24.346	3.439	0.281
Nasal Height			2.154	10,774	2.500	0.400
Nasal Breadth			0.028	7.949	142.857	0.007
Cephalic Index			3.311	24.663	3.723	0.269
Nasal Index .			10.779	58.777	2.727	0.369
Morphological Fac	ial In	dex	1.199	19.192	8.000	0.125
Minimum Frontal	Diam	eter	0.892	20.322	11.389	0.087
Bigonial Diameter	r		5.170	33.014	3.187	0.313
Physiological Face	e Leng	th	14.749	47.594	1.614	0.610
Internal Ocular B	readth	ì	0.392	7.873	10.040	0.090
External Ocular I	Breadt	h	10.876	17.970	0.827	1.210
Height of Head			2.086	46.529	11.161	0.089
Ear Length .	•	•	3.985	20.442	2.566	0.390

- 19. Looking at Col. (5) in Table VI we notice that the variations of Nasal Breadth ('007), Minimum Frontal Diameter ('087), Height of the Head ('089) and Internal Ocular Breadth ('090) are extremely small, and therefore these characters remain practically constant for the whole family.
- Morphological Facial Index (125), Head Length (191), Morphological Face Length (260), Cephalic Index (269), Bizygomatic Breadth (281), Bigonial Diameter (313), and Head Breadth (333) are moderately variable, while the greatest variations within the family occur in Nasal Index (367), Ear-Length (390), Nasal Height (400), Physiological Face Length (610), and to a much more pronounced degree in the External Ocular Breadth (1210).
- 20. A glance at Table I(B) shows that among the Chinese there is quite a gradual decrease in the External Ocular Breadth (from 95.6 mm. to 88.1 mm.), Physiological Face Length (from 192.1 to 183.6), Ear-length (from 64.1 to 59.6), Morphological Face Length (from 117.8 to 112.5), Bigonial Diameter (from 109.2 to 105.3) and Head Length (from 187.8 to 185.2), as we pass from the northern to the southern provinces. Koreans fall in line with the Chinese for every character except the Head Length, while Manchus differ only in Head Length and Physiological Face Length.
- 21. The present study indicates therefore that Nasal Breadth, Minimum Frontal Diameter, Height of the Head, and Internal Ocular Breadth are constant or family characteristics for the provincial samples discussed here, while the variation within the family is most strongly marked in External Ocular Breadth, Ear-length, Morphological Face Length, Bigonial Diameter, Physiological Face Length and Head Length.