动差、新动差、乘积动差及其相互间关系

動差，新動差，乘積動差及其相互間關係

汪厥明 著

Moments, Cumulants, Product-moments and Relations *inter se*.

By. C. M. Wang

Published by the

Chueh-Ming Biometry Laboratory

厥明生物統計研究室出版

1 9 4 7

動差，新動差，乘積動差及其相互間關係

Moments, Cumulants, Product-moments and Relations *inter se.*

每冊定價臺幣三百五十元 （郵資另加）

著　　者　汪　厥　明

發 行 所　厥明生物統計研究室

臺灣臺北市國立臺灣大學農學院轉

印 刷 所　臺灣新生報社新生印刷廠

臺北市西園街一三五五號

中華民國三十六年六月初版

SYNOPSIS

MOMENTS, CUMULANTS, PRODUCT-MOMENTS AND RELATIONS *inter se*

by Chueh-Ming Wang

I. The parameters of population are the objects of biometrical study, but they are frequently imarginary and abstract truth so that we could not show them as such concretely. In such cases the only thing we can do is to take from parent population the random samples from which we estimate the correspondent statistics instead of parameters.

II. The methods of sampling are divided into:

(1) sampling with replacement

and (2) sampling without replacement.

The former, in gerenal, is called the simple sampling. As a result of simple sampling an infinite population can be formed of a finite one. But under the circumstances of sampling without replacement an infinite population can arise from the infinite one only.

III. In the sense of statistical probability the most probable value of a statistics is its expectation. A statistics may yield one or more than one value for its expectation according as whether its parent population is finite or not and the size of population in turn, may vary as sampling methods. Therefore, on considering the expectation of a statistics we must refer to the sampling procedures involved.

IV. The axiom Prof. Fisher, R. A. and his followers believe in seems to be that every population might be assumed to be infinite as a result of simple sampling. Indeed, under this axiom some theorems worked out by him *et al* are at best and can admit, perhaps, no further improvement.

V. Although the writer refers to both the two cases of sampling with and without replacement when considering samples involved, still, he feels with Prof. Fisher and is convinced that when assuming the infinity of population as a result of simple sampling there is frequently great convenience but no absurdity, especially, in dealing with k-statistics.

VI. Both the moments and the product-moments here involved are divided into two categories. For example, the moments are divided into:

(1) μ'_r, the moments of rth order about zero:

$$\mu'_r=\frac{1}{N}\sum_{i=1}^{m}n_i\,x_i^r=\frac{1}{N}\sum_{i=1}^{m}x_i^r$$

where x_i is a variate measured from zero, n_i the frequency of the peculiar value of x_i, r its exponent and N the size of population.

and (2) μ_r, the moment of rth order about arithmetic mean:

$$\mu_r=\frac{1}{N}\sum_{i=1}^{m}n_i\,(x_i-\mu)^r=\frac{1}{N}\sum_{i=1}^{m}(x_i-\mu)^r=\frac{1}{N}\sum_{i=1}^{m}n_i\,x_i^r=\frac{1}{N}\sum_{i=1}^{m}x_i^r$$

where the variate measured from population mean (μ) each equal to $(x_i-\mu)$ and the other symbols are just the same as defined in (1).

According to (1) when $r=1$, μ'_1 is the first order moment about zero and therefore, the population mean (μ) and according to (2) when $r=2$, μ_2 is the second order moment about population mean and thus, equal to the variance of distribution.

VII. The relations between μ'_r and μ_r can be shown simply in the following expressions:

(1) $\mu'_r={}^rC_0\,\mu_r+{}^rC_1\,\mu_{r-1}\,\mu'_1+{}^rC_2\,\mu_{r-2}\,\mu'^2_1+\cdots+{}^rC_{r-1}\,\mu_1\,\mu'^{r-1}_1+{}^rC_r\,\mu'^r_1$

and (2) $\mu_r={}^rC_0\,\mu'_r-{}^rC_1\,\mu'_{r-1}\,\mu'_1+{}^rC_2\,\mu'_{r-2}\,\mu'^2_1-{}^rC_3\,\mu'_{r-3}\,\mu'^3_1+\cdots+(-1)^r(1-r)\mu'^r_1$

VIII. Population mean of the product of variates such as $x_i^{r_1}$, $x_j^{r_2}$, …, and $x_l^{r_h}$ is the product-moment of $r_1, r_2, \ldots, r_h$th order and symbolised by

(1) $\mu'_{r_1.r_2.\ldots r_h}$. the product-moment about zero of $r_1, r_2, \ldots, r_h$th order:

$$\mu'_{r1,\,r2,\,\ldots,\,rh}=E\,(\mathrm{x}_i^{r1}\,\mathrm{x}_j^{r2}\,\ldots\,\mathrm{x}_l^{rh})=\frac{\sum\limits_{i\neq j}\sum\limits_{\neq k}\cdots\sum\limits_{\ldots\neq l}\mathrm{x}_i^{r1}\,\mathrm{x}_j^{r2}\cdots\mathrm{x}_l^{rh}}{N(N-1)(N-2)\cdots(N-h+1)}$$

where x_i, x_j, $\cdots$, x_l are the variates of the same population measured from zero and $E\,(\mathrm{x}_i^{r1}\,\mathrm{x}_j^{r2}\ldots\mathrm{x}_l^{rh})$ is the expectation of their product

and (2) $\mu_{r1,\,r2,\,\ldots,\,rh}$ the product-moment about mean of r_1, r_2, $\cdots$, r_hth order:

$$\mu_{r1,\,r2,\,\ldots,\,rh}=E\,(x_i^{r1}\,x_j^{r2}\ldots x_l^{rh})=\frac{\sum\limits_{i\neq j}\sum\limits_{\neq k}\cdots\sum\limits_{\ldots\neq l}x_i^{r1}\,x_j^{r2}\cdots x_l^{rh}}{N(N-1)(N-2)\cdots(N-h+1)}$$

where x_i, x_j, $\cdots$, x_l are the variates of the same population measured from mean and $E\,(x_i^{r1}, x_j^{r2}, \cdots, x_l^{rh})$ the expectation of their product.

IX. The relations between moments and product-moments of the same population are considered here in both the case of sampling with and without replacement:

(1) In case of variates measured from O and sampled without replacement:

(i) $$\mu'_{r1,\,r2}=\frac{N\mu'_{r1}\,\mu'_{r2}-\mu'_{r1+r2}}{N-1}$$

(ii) $$\mu'_{r1,\,r2,\,r3}=\frac{1}{(N-1)(N-2)}\Big[N^2\mu'_{r1}\,\mu'_{r2}\,\mu'_{r3}-N\,(\mu'_{r1+r2}\,\mu'_{r3}+\mu'_{r1+r3}\,\mu'_{r2}+\mu'_{r2+r3}\,\mu'_{r1})+2\,\mu'_{r1+r2+r3}\Big]$$

(iii) $$\begin{aligned}\mu'_{r1,\,r2,\,r3,\,r4}=\frac{1}{(N-1)(N-2)(N-3)}\Big[&N^3\mu'_{r1}\,\mu'_{r2}\,\mu'_{r3}\,\mu'_{r4}-N^2\,(\mu'_{r1+r2}\,\mu'_{r3}\,\mu'_{r4}\\&+\mu'_{r1+r3}\,\mu'_{r2}\,\mu'_{r4}+\mu'_{r1+r4}\,\mu'_{r2}\,\mu'_{r3}+\mu'_{r2+r3}\,\mu'_{r1}\,\mu'_{r4}+\mu'_{r2+r4}\,\mu'_{r1}\,\mu'_{r3}+\mu'_{r3+r4}\,\mu'_{r1}\,\mu'_{r2})\\&+2N\,(\mu'_{r1+r2+r3}\,\mu'_{r4}+\mu'_{r1+r2+r4}\,\mu'_{r3}+\mu'_{r1+r3+r4}\,\mu'_{r2}+\mu'_{r2+r3+r4}\,\mu'_{r1})\\&+N\,(\mu'_{r1+r2}\,\mu'_{r3+r4}+\mu'_{r1+r3}\,\mu'_{r2+r4}+\mu'_{r1+r4}\,\mu'_{r2+r3})-6\,\mu'_{r1+r2+r3+r4}\Big]\end{aligned}$$

etc.

and (2) In case of variates measured from O and sampled with replacement: Since N, the size of population tends to infinity, relations between $\mu'_{r1,\,r2,\,\ldots,\,rh}$ and μ'_r can be shown simply in such general expression as follows:

$$\mu'_{r1,\, r2,\, \dots,\, rh} = \mu'_{r1}, \mu'_{r2}, \dots, \mu'_{rh}\,.$$

If variates are measured from mean, we substitute μ_r for μ'_r in the expressions above. It is worthy of note that in case of sampling with replacement, if any one of r's equals unity, the relevant product-moment vanishes owing to such circumstances as follows:

$$\mu_{1,\, r2,\, r3,\, \dots,\, rh} = \mu_{r1,\, 1,\, r3,\, \dots,\, rh} = \cdots = \mu_1\, \mu_{r2}\, \mu_{r3} \cdots \mu_{rh} = \mu_{r1}\, \mu_1\, \mu_{r3} \cdots \mu_{rh} \cdots$$
$$= \mu_{r1}\, \mu_{r2}\, \mu_{r3} \cdots \mu_1 = 0.$$

X. Let us now turn to the relations among m-statistics* such as m_r, m'_r $m_{r1,\, r2,\, \dots,\, rh}$ and $m'_{r1,\, r2,\, \dots,\, rh}$ which are like those between moments and product-moments of population and can be brought out by substituting m for μ and n for N in case of sampling without replacement. In this case m_1 also equals zero.

XI. Moments or product-moments through which the observation and expectation are combined are called the combined moments or the combined product-moments. c_r, the combined moments and $c_{r1,\, r2,\, \dots,\, rh}$, the combined product-moments are essentially related to random errors (ε) so that they could be defined as follows:

(1) c_r:

$$c_r = \frac{1}{n} \sum_{i=1}^{n} (\dot{x}_i - \mu'_1)^r = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^r\,, \qquad (\varepsilon_i = \dot{x}_i - \mu'_1)$$

and (2) $c_{r1,\, r2,\, \dots,\, rh}$:

$$c_{r1,\, r2,\, \dots,\, rh} = \frac{1}{n(n-1)(n-2)\cdots(n-h+1)} \sum_{i \neq} \sum_{j \neq} \cdots \sum_{\neq l} (\dot{x}_i - \mu)^{r1} (\dot{x}_i - \mu)^{r2} \cdots (\dot{x}_i - \mu)^{rh}$$
$$= \frac{1}{n(n-1)\cdots(n-h+1)} \sum_{i \neq} \sum_{j \neq} \cdots \sum_{\neq l} \varepsilon_i^{r1}\, \varepsilon_j^{r2} \cdots \varepsilon_l^{rh}$$

where $\dot{x}_i$ is a variate from a sample of n individuals.

(3) Relations between c_r and $c_{r1,\, r2,\, \dots,\, rh}$: These relations are like those between μ_r and $\mu_{r1,\, r2,\, \dots,\, rh}$, but c stands for μ.

---

*) $m'_r = \frac{1}{n}\sum_{i=1}^{n} \dot{x}_i^r$, $m_r = \frac{1}{n}\sum_{i=1}^{n} (\dot{x}_i - \bar{x})^r = \frac{1}{n}\sum_{i=1}^{n} \dot{x}_i^r$, etc.

XII. The combined moments can be calculated and analysed by means of the following formulae:

(1) Calculating formula:

$$c_r = m'_r - r m'_{r-1}\mu'_1 + \frac{r(r-1)}{2!} m'_{r-2}\mu'^2_1 - \cdots (-1)^{r-1} m'_1 \mu'^{r-1}_1 + (-1)^r \mu'^r_1$$

and (2) Analysing formula:

$$c_r = m_r + r m_{r-1}\bar{\varepsilon}_s + \frac{r(r-1)}{2!} m_{r-2}\bar{\varepsilon}^2_s + \cdots + \bar{\varepsilon}^r_s .$$

where $\bar{\varepsilon}_s = \bar{\mathrm{x}} - E\mathrm{x} = m'_1 - \mu'_1 = \frac{1}{u}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_n) = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i$

From the above it follows that c_r may be expressed in two ways, m'_r and μ'_1 being involved in one of them and m_r and $\bar{\varepsilon}$ in the another, and that the variation of c_r is due to two sources, viz., variation within and between samples. The analysis of $c_{r1,\, r2,\, \ldots,\, rh}$ may be exemplified by that of $c_{r1,\, r2}$.:

$$c_{r1,\, r2} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j} \{[\xi^{r1}_i + r_1 \xi^{r1-1}_i \bar{\varepsilon}_s + \cdots + \bar{\varepsilon}^{r1}_s][\xi^{r2}_j + r_2 \xi^{r2-1}_j + \cdots + \bar{\varepsilon}^{r2}_s]\},$$

where $\xi_i = \dot{\mathrm{x}}_i - \bar{\mathrm{x}}$ and $\xi_j = \dot{\mathrm{x}}_j - \bar{\mathrm{x}}$.

XIII. The expectation of statistics may be regarded as a population value and thus, a combined parameter connecting sample and its parent population. The expectations of various statistics are enumerated below:

(1) m'_r: the expectation of m'_r in general, is μ'_r, no matter whatever the size of population may be,

and (2) m_r: The expectation of m_r varies as the sampling methods and highness of its order:

(i) Sampling without replacement:

a) $Em_0 = 1$, b) $Em_1 = 0$, c) $Em_2 = \frac{N}{N-1} \cdot \frac{n-1}{n} \mu_2$,

d) $Em_3 = \frac{N^2}{(N-1)(N-2)} \cdot \frac{(n-1)(n-2)}{n^2} \mu_3$

and e) $$Em_4=\frac{N}{(N-1)(N-2)(N-3)}\cdot\frac{(n-1)(n-2)(n-3)}{n}$$
$$\times[(N^2-2N+3)\mu_4-3(2N-3)\mu_2^2]$$
$$+\frac{(n-1)(2n-3)}{2}\cdot\frac{N}{N-1}[\mu_4+3\mu_2^2]$$

where N and n are the size of population and sample respectively.

(ii) Sampling with replacemant: In this case all the expectations are different from and simpler than the above except when $r=0,1$.

a) $Em_2=\frac{n-1}{n}\mu_2$, b) $Em_3=\frac{(n-1)(n-2)}{n^2}\mu_3$,

c) $$Em_4=\frac{n-1}{n^3}[(n^2-3n+3)\mu_4+3(2n-3)\mu_2^2]$$

and d) $$Em_2^2=\frac{n-1}{n^3}[(n-1)\mu_4+(n^2-2n+3)\mu_2^2].$$

From the above we see that c) and d) are two simultaneous equations with the two unknowns μ_4 and μ_2^2 and that if the values of μ_4 and μ_2^2 acquired are substituted in expression of κ_4, the following formula will be obtained:

$$\kappa_4=\frac{n^2}{(n-1)(n-2)(n-3)}[(n+1)Em_4-3(n-1)Em_2^2]$$

We also see that k-statistics, the estimates of κ_r must be related to m_2^2 and m_4 in such a way as follows:

$$k_4=\frac{n^2}{(n-1)(n-2)(n-3)}[(n+1)m_4-3(n-1)m_2^2].$$

This is an instance that illustrates how the calculating formulae of k-statistics are derived. In either case of sampling the expectation of m_r can be calculated by means of either of the following formulae, but that of f) is more convenient.

e) $$Em_r=E[m'_r-rm'_{r-1}m'_1+\frac{r(r-1)}{2!}m'_{r-2}m_1'^2-\cdots+(-1)^r(1-r)m_1'^r]$$

f) $$Em_r=E[c_r-rc_{r-1}c_1+\frac{r(r-1)}{2!}c_{r-2}c_1^2-\cdots+(-1)^r(1-r)c_1^r]$$

(3) $m'_{r1,\, r2,\, \ldots,\, rh}$ and $m_{r1,\, r2,\, \ldots,\, rh}$: The expectation of $m'_{r1,\, r2,\, \ldots,\, rh}$ is always $\mu'_{r1,\, r2,\, \ldots,\, rh}$, but that of $m_{r1,\, r2,\, \ldots,\, rh}$, must be calculated each by means of each correspondent formula.

(4) c_r: The expectation of c_r—the combined moment—varies in accordance with sampling methods, but all the combined moments can be analysed in such a way as follows:

$$Ec_r = E\left(m_r + r m_{r-1}\bar{\varepsilon}_s + \frac{r(r-1)}{2!} m_{r-2}\bar{\varepsilon}_s^2 + \cdots + \bar{\varepsilon}_s^r\right) = \mu_r$$

(5) $c_{r1,\, r2,\, \ldots,\, rh}$: The following expression is available for the expression of $c_{r1,\, r2,\, \ldots,\, rh}$ in either case of sampling:

$$Ec_{r1,\, r2,\, \ldots,\, rh} = \mu_{r1,\, r2,\, \ldots,\, rh}$$

where $\mu_{r1,\, r2,\, \ldots,\, rh}$ varies according to the sampling procedures.

XIV$_a$. k-statistics and k-parameters are established by Professor Fisher, R.A., the latter being called the cumulants. Under the presumption of infinite population these parameters are very useful. They are qualified as follows:

(1) k-Statistics: These statistics consist of S_r, s_r and n, the size of sample, s_r and S_r being defined by professor Fisher as follows:

$$s_r = \sum_{i=1}^{n} \mathrm{x}_i^r = n m'_r, \quad S_r = \sum_{i=1}^{n} (\mathrm{x}_i - \bar{\mathrm{x}})^r = \sum_{i=1}^{n} \dot{x}_i^r = n m_r$$

Thus, (i) $k_1 = \dfrac{s_1}{n} = m'_1 = \bar{\mathrm{x}}$,

(ii) $k_2 = \dfrac{1}{n-1} S_2 = \dfrac{n}{n-1} m_2$,

(iii) $k_3 = \dfrac{n}{(n-1)(n-2)} S_3 = \dfrac{n^2}{(n-1)(n-2)} m_3$,

(iv) $k_4 = \dfrac{n}{(n-1)(n-2)(n-3)} \left[(n+1)S_4 - \dfrac{3(n-1)}{n} S_2^2\right]$

$= \dfrac{n^2}{(n-1)(n-2)(n-3)} \left[(n+1) m_4 - 3(n-1) m_2^2\right]$

and (v) $k_4 = \dfrac{n^2}{(n-1)(n-2)(n-3)(n-4)} \left[(n+5) S_5 - \dfrac{10(n-1)}{n} S_3 S_2\right]$

$$= \frac{n^3}{(n-1)(n-2)(n-3)(n-4)} [(n+5)m_5 - 10(n-1)m_3 m_2]$$

Ek_r, the expectation of k_r of course, is κ_r, the cumulant.

(2) Relations between κ_r and μ_r: From knowledge of the expectation of k-statistics we see how cumulants and moments are related to each other, viz.,

(i) $\kappa_1 = Ek_1 = \mu_1' = \mu$, (ii) $\kappa_2 = Ek_2 = \mu_2$,

(iii) $\kappa_3 = Ek_3 = \mu_3$, (iv) $\kappa_4 = Ek_4 = \mu_4 - 3\mu_2^2$

and (v) $\kappa_5 = Ek_5 = \mu_5 - 10\mu_3\mu_2$,

(3) Relations between κ_r and λ_r: λ_r's are Thiele's semi-invariants. Because in calculating λ_r he used the statistics m_r instead of μ_r, the expectations of these parameters—it would be rather better to say "a kind of Statistics"—are different from cumulants except $E\lambda_1 = \kappa_1$, viz.,

(i) $E\lambda_1 = \kappa_1$, (ii) $E\lambda_2 = \frac{n-1}{n}\kappa_2$, (iii) $E\lambda_3 = \frac{(n-1)(n-2)}{n^2}\kappa_3$,

(iv) $E\lambda_4 = \frac{n-1}{n^3}[(n-6n+6)\kappa_4 - 6n\kappa_2^2]$

and (v) $E\lambda_5 = \frac{(n-1)(n-2)}{n^4}[(n-12n+12)\kappa_5 - 60n\kappa_3\kappa_2]$

From the above it is evident that semi-invariants and cumulants are not thoroughly identical.

XIV$_b$. In addition to the infinity of size of population, randomness and independence are two important proporties of error. Such things agree with Professor Fisher's conception about the variate $\dot{x}_i$ and therefore, both the combined and non-combined moments and product-moments can be interpreted in terms of error, that is to say;

(1) Relations between error and combined moments: Such relations can be brought out by following expressions, viz.,

$$\dot{x}_i = \mu_1' + \varepsilon_i, \quad \varepsilon_i = \dot{x}_i - \mu_1'; \quad \bar{x} = \mu_1' + \bar{\varepsilon}_s, \quad \bar{\varepsilon}_s = \bar{x} - \mu_1'.$$

hence, (i) $c_r = \frac{1}{n}\sum_{i=1}^{n}(\dot{x}_i - \mu'_1)^r = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^r$

and (ii) $c_{r1,\,r2,\,\ldots,\,rh} = \frac{1}{n(n-1)(n-2)\cdots(n-k+1)}\sum_{i=1}^{n}\sum_{i\neq j\neq}\cdots\sum_{\neq l}\varepsilon_i^{r1}\,\varepsilon_j^{r2}\cdots\varepsilon_l^{rh}$

(2) Relations between error and μ_r: An understanding of these relations is very important on applying them to the research practice and they can be shown as follows:

(i) $E\varepsilon_i = \mu_r$,

and (ii) $E\varepsilon_i^{r1}\,\varepsilon_j^{r2}\cdots\varepsilon_l^{rh} = \mu_{r1}\,\mu_{r2}\cdots\mu_{rh}$

and whereby we see that:

(iii) $E\varepsilon_i = \mu_1 = 0$

and (iv) $E\varepsilon_1^{r1}\,\varepsilon_j^{r2}\cdots\varepsilon_l^{rh} = \mu_1\,\mu_{r2}\cdots\mu_{rh} = \cdots = \mu_{r1}\,\mu_{r2}\cdots\mu_1 = 0$

XV. Since what we can practically deal with is the sample mean that fluctuates sample by sample, the variance [$V(\bar{x})$] and standard deviation [S.D. $(\bar{x})$] of mean are very important for the purposes of practice, but they vary with the sampling methods in such a way as follows:

(1) In case of sampling without replacement:

(i) $V(\bar{x}) = \frac{N-n}{N-1}\cdot\frac{\mu^2}{n} = \frac{N-n}{N-1}\cdot\frac{k_2}{n}$

and (ii) $S.D.(x) = \sqrt{\frac{N-n}{N-1}\cdot\frac{\mu_2}{n}} = \sqrt{\frac{N-n}{N-1}\cdot\frac{k_2}{n}}$

(2) In case of simple sampling:

(i) $V(\bar{x}) = \frac{\mu_2}{n}$,

and (ii) $S.D.(\bar{x}) = \sqrt{\frac{\mu_2}{n}}$

The variance of mean also can be brought out by means of $E(c_1 - Ec_1)^2$.

XVI. Because both the mean square and standard error are estimated from sample, they vary about their true values. In the case of simple sampling these variances due to random variation are calculated in such a way as follows:

(1) Variance for mean square$=E(k_2-Ek_2)^2=\frac{k_2}{n}+\frac{2k_2}{n-1}$

and (2) Variance for standard error$=\frac{k_4}{4nk_2}+\frac{k_2}{2(n-1)}$

Taking the square roots of (1) and (2) we obtain the standard deviations for mean and standard error respectively. In the case of small sample - that is to say n is small—we would rather call both of them the standard error than the standard deviation.

XVII. g_1 and g_2, the estimates of γ_1 and γ_2, the parameters related to skewness and kurtosis and their variances are such as follows:

(1) Estimating γ_1 and γ_2:

(i) $g_1=\frac{k_3}{k_2^{3/2}}$

and (ii) $g_2=\frac{k_4}{k_2^2}$.

and (2) Variance of g_1 and g_2:

(i) Variance for $g_1=E(g_1-Eg_1)^2=\frac{E(k_3^2)}{E(k_2^3)}-\frac{k_3^2}{k_2^3}=\frac{n^2(n-1)^2}{n(n-1)(n-2)}$

$$\times\frac{[(n-1)(n-2)k_6+9n(n-2)k_4k_2+n(n-2)(n+8)k_3^2+6n^2k_2^3]}{[(n-1)(n-2)k_6+3n(n-1)(n+3)k_4k_2+4(n-2)k_3^2+n(n+1)(n+3)k_2^3]}-\frac{k_3^2}{k_2^3}$$

and (ii) Variance for $g_2=E(g_2-Eg_2)^2=\frac{E(k_4^2)}{E(k_2^4)}-\frac{k_4^2}{k_2^4}$

The expression of variance for g_2 is so lengthy that we are obliged to omit it. In this connection, readers are requested to refer to the text.

XVIII. If of N individuals of parent population of binomial distribution there are M with attribute A, then its chance of success will be $p=\frac{M}{N}$, by means of which the important parameters and statistics of this distribution can be derived.

(1) Sampling without replacement: If the relative frequency of m individuals with attribute A and $(n-m)$ with B in a sample of n individuals is P_i', then it will be such as follows:

$$P_i'=\frac{n(n-1)\cdots(m+2)(m+1)}{1\cdot 2\cdot 3\cdots(n-m+1)(n-m)}$$

$$\times\frac{M(M-1)(M-2)\cdots(M-m+1)(N-M)(N-M-1)(N-M-2)\cdots(N-M-n+m+1)}{N(N-1)(N-2)\cdots(N-n+1)}$$

where M is the number of individuals with attribute A in a population of N.

(i) Calculating μ_r':

a) $\mu_1'=np$, b) $\mu_2'=np[(n-1)p_1+1]$,

c) $\mu_3'=np[(n-1)(n-2)p_1p_2+3(n-1)p_1+1]$

d) $\mu_4'=np[(n-1)(n-2)(n-3)p_1p_2p_3+6(n-1)(n-2)p_1p_2$

$+7(n-1)p_1+1]$

and (ii) Calculating μ_r:

a) $\mu_1=\mu_1'-\mu_1'=0$, b) $\mu_2=np[(n-1)(p_1-p)+q]$

c) $\mu_3=np[n^2(p_1p_2+2p-3pp_1+3np_1(p-p_2)+3n(p_1-p)+2p_1p_2-3p_1+1]$

and d) $\mu_4=np[n^3(p_1p_2p_3+6p^2p_1-pp_1p_2-3p^3)+6n^2(2pp_1p_2-p_1p_2p_3-p^2p_1)$

$+np_1p_2(11p_3-8p)-6p_1p_2p_3+6n^2(p_1p_2+p^2-2pp_1)-6np_1(3p-2p)$

$+12p_1p_2+n(7p_1-4p)-7p_1+1]$

Where $q=1-p$, $p_1=\frac{M-1}{N-1}$, $p=\frac{M-2}{N-1}$ and $p_3=\frac{M-3}{N-3}$,

If N and M are very large, then p_1, p_2, and p_3 are all practically equal to p and the expressions of b), c) and d) turn simple, viz.,

b) $\mu_2\simeq npq$, c) $\mu_3\simeq npq(q-p)$

and d) $\mu_4\simeq 3n^2p^2q^2+npq(1-6pq)$.

(2) Sampling with replacement: In this case the general expression of P_i, the relative frequency is such as follows:

$$P_i=\frac{n!}{m!(n-m)!}p^mq^{n-m}$$

where p, the chance of success always holds good.

(i) Calculating μ'_r:

a) $\mu'_1 = np$, b) $\begin{cases} \text{for number of success} \cdots\cdots \mu'_2 = np[(n-1)p+1] \\ \text{for proportion of success} \cdots \mu'_2 = \frac{p}{n}[(n-1)p+1] \end{cases}$

c) $\mu'_3 = np[((n-1)(n-2)p^2+3(n-1)p+1]$

and d) $\mu'_4 = np[(n-1)(n-2)(n-3)p^3+6(n-1)(n-2)p^2+(n-1)p+1]$

and (ii) Calculating μ_r:

a) $\mu_1 = 0$, b) $\begin{cases} \text{for number of success} \cdots\cdots \mu_2 = npq \\ \text{for proportion of success} \cdots\cdots \mu_2 = \frac{pq}{n} \end{cases}$

c) $\mu_3 = npq(q-p)$

and d) $\mu_4 = 3n^2 p^2 q^2 + npq(1-6pq)$

XIX. If of n individuals in a sample there are m with attribute A, then the chance of success of A will be $p'\left(=\frac{m}{n}\right)$ that is the estimate of population value, p. When substituting p' and $q'(=1-p')$ for p and q in the expressions of μ'_r and μ_r, we obtain those of m'_r and m_r. In the case of sampling without replacement p'_1, p'_2 and p'_3 are substituted for p_1, p_2 and p_3 where p'_1, p'_2 and p'_3 are equal to $\frac{m-1}{n-1}$, $\frac{m-2}{n-1}$ and $\frac{m-3}{n-3}$ respectively. Let us now turn to the relations between Em_2 and μ_2 which are such as follows:

(1) Sampling without replacement:

(i) For number of success: $\mu_2 = np[(n-1)(p_1-p)+q]$

$$= \frac{N-1}{N} \cdot \frac{n}{n-1} Enp'[(n-1)(p'_1-p')+q']$$

and (ii) For proportion of success: $\mu_2 = \frac{p}{n}[(n-1)(p_1-p)+q]$

$$= \frac{N-1}{N} \cdot \frac{n}{n-1} E\frac{p'}{n}[(n-1)(p'_1-p')+q']$$

and (2) Sampling with replacement:

(i) For number of success: $\mu_2 = npq = \frac{n}{n-1} Enp'q'$

and (ii) For proportion of success: $\mu_2 = \frac{pq}{n} = \frac{n}{n-1} E\frac{p'q'}{n}$

XX. As to cumulants and k-statistics, only the parameters and k-statistics for the number of success are shown here with the exception of those of second order which are very important in practice of research works. As a matter of course, here only the sampling with replacement is concerned in it.

(1) Calculating k_r:

(i) $k_1 = np'$, (ii) $\begin{cases} \text{For number of success: } k_2 = \frac{n}{n-1} np'q', \\ \text{For proportion of success: } k_2 = \frac{n}{n-1} \cdot \frac{p'q'}{n} \end{cases}$

(iii) $k_3 = \frac{n^2}{(n-1)(n-2)} np'q'(q'-p')$

and (iv) $k_4 = \frac{n^2}{(n-1)(n-2)(n-3)} [np'q' + np'q'(1-6p'q')]$

and (2) Calculating k_r:

(i) $k_1 = np$, (ii) $\begin{cases} \text{For number of success} \ldots\ldots k_2 = npq \\ \text{For proportion of success} \ldots\ldots k_2 = \frac{p \cdot q}{n} \end{cases}$

(iii) $k_3 = npq(q-p)$

and (iv) $k_4 = npq(1-6pq)$.

XXI. The γ_1 and γ_2 of binomial distribution in a sample of n individuals from the infinite population may be expressed in terms of p and q as follows:

(1) $\gamma_1 = \frac{q-p}{\sqrt{npq}}$,

nd (2) $\gamma_2 = \frac{1-6pq}{npq}$.

From the above we see that when n is very large both γ_1 and γ_2 tend to be zero, viz., the binomial distribution inclines to normality.

XXII. The error—in other words, the random error—is normally distributed and fluctuate between $\pm\infty$, so as its population is infinite and its distribution can be shown in the following forms:

(1) Generating Function: It can be expressed in three forms, viz.,

(i) $df=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(X-\mu)^2}{2\sigma^2}}dx$, (x measured from O as origin)

(ii) $df=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}dx$, ($x$ measured from mean μ as origin)

(iii) $df=\frac{1}{\sqrt{2\pi}}e^{-\frac{w^2}{2}}d\omega$, ($\omega=\frac{x}{\sigma}$, $dx=\sigma d\omega$.)

From the above it is evident that the normal curve is symmetrical about the ordinate located at mean as axis so that all the moments of odd numbered order disappear and that the total probability of variate obeying the normal law of error must equal unity.

(2) Calculating μ_r: The moments about mean of r th order are very important for the purposes of practice and can be calculated directly in such a way as below:

$$\mu_r=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty}x^r e^{-\frac{x^2}{2\sigma^2}}dx=\frac{2}{\sigma\sqrt{2\pi}}\int_0^{+\infty}x^r e^{-\frac{x^2}{2\sigma^2}}dx=\frac{r!}{2^{\frac{r}{2}}\left(\frac{r}{2}\right)!}\sigma^r$$

From the properties of expression on the right side it is obvious at once that in normal distribution all moments about mean of odd numbered order vanish and that those of even numbered order could be brought out alternatively by means of the following formula:

$$\mu_r=1\cdot3\cdot5\cdots(r-3)(r-1)\sigma^r$$

When substituting $r=0,2,4,6,8,\cdots$ each by each in the formula above, we obtain: