河流水文参数的統計計算技术

一

苏联 K. M. 伊万諾夫斯卡婭著
徐在庸譯

电力工業出版社

前　言

本書是实际工作中必备的参考書，它通俗地介紹了水文統計参数的計算技术，从而可以确定各种現象間的規律和关系，并且在介紹时还利用了变差統計学。

水文現象是由各种不同的、極复杂的物理現象总体所組成。

在水文現象循环(年循环、年內循环、多年循环等)的各种組合中，根据过去多年观測資料来作水文預报的問題，以及水文資料与气象資料、气候資料等相关問題，是現代水文学中很重要的一个方面。

在水文学中，与在其他許多自然科学中一样，規律性并不表示为起始条件与有关現象之間的准确数学关系，或者还可以这样說，規律性并不表示为因果关系。統計計算在很大的或然程度上能够闡明这种关系，而在許多情形下，統計計算可能是解决若干水文問題和水利經济問題以及作出有很高或然程度的水文預报的一种主要手段。

統計計算分为兩部分：数理統計学（就这个詞的狹义来說）与变差統計学。

在数理統計学中是根据概率理論来研究統計变数的一般規律，并得出整理統計变数的各种公式和各种方法的論据。

变差統計学是一門純粹的实用科目，主要是帮助应用已拟定的各种公式，并將主要注意力放在計算技术上；其中給出了計算工作的方法与标准格式，采用这些計算方法与标准格式，就可以

我們認为

一种將实际資料加以

系統化的深入的和扩大的方法时，具有重大的意义，但这絕不意味着統計計算是唯一的万能的方法，也不意味着把統計計算与成因法对立起来。自然界的現象是如此的复杂，以至于要想把它們化簡为数学問題就会是一个錯誤；在全部工作过程中，統計研究都应該伴随有全面的成因分析。

在本書中，列有数理統計学所求到的許多有根据的一般結論，然而，本書的重心乃是变差統計学。在 A. K. 米特罗波尔斯基教授［参考文献 11］Ю.Л. 波莫尔斯基教授［参考文献 15］，以及 B.И. 罗曼諾夫斯基教授［参考文献 16］的著作中，都以最通俗的形式叙述了变差統計法。在編写本書时，这些著作被用为叙述計算方法、計算公式和用算例来說明的格式等部分的基础。

正文分为三章。第一章說明統計計算的一般概念和統計計算的各种公式。第二章中，叙述計算統計参数的技术，并通过按标准計算表格演算的各个数字例題，来說明各种計算的方法。第三章中，列有作者根据統計計算法則所推得的各公式与諾謨圖，这些公式和諾謨圖可以使进行水文和水利經济調查以及設計时所用的若干計算大大地簡化和加速，例如按年逕流量的变差系数和偏差系数計算值来决定平水流量的相应値；將年流量的关系方程换算为逕流的关系方程，以及將逕流的关系方程换算为年流量的关系方程等工作。

变差統計学可以得出若干計算参数的方法，这些方法的应用按系列長短、計算目的等而定。在第二章中，叙述了水文学中通常所用的若干种方法：直接計算法，其另一个名称叫作“最小二乘”法（馬尔科夫法）；直接計算的复杂法——“水文学”法；以及作者所建議采用的“总和”法。

“取整”法之所以列入本書，專門是为了便于理解“总和”法的計算公式，“取整”法好像是“总和”法的先导。

直到現在为止，在水文实踐中采用“总和”法的还比較少見，虽然“总和”法有一些很重要的优点；此外，A.K.米特罗波尔斯基、Ю.Л.波莫尔斯基等都認为“总和”法是一种最簡單和最迅速的方法，在工作过程中可以很容易地进行校核，并且能保証和較繁杂方法具有同样的精度。在苏联，首先是由П.А.叶菲莫維奇工程师將总和法应用于水文实踐中，在他的指导下，本書作者曾完成了莫斯科河水文概貌的一切計算，以及莫斯科运河設計中的一部分水利經济計算[参考文献5]。

这里，必須批判地談到直接計算的复杂法——“水文学”法，在現代水文实踐中所采用的就是这种方法。

其所以复杂就在于換算，例如，將河川逕流量換算成模量系数。这样对水文統計参数的計算結果并無影响，但却使計算过程变得更加繁重。在最近一些水文著作中(如Б.А.阿波洛夫，[参考文献1]、Е.В.布利茲尼亞克[参考文献3]、А.В.奥吉耶夫斯基[参考文献13]、Б.В.波利亞科夫[参考文献14])，直接計算的簡單法，或所謂最小二乘法(馬尔科夫法)已开始代替直接計算的复杂法——水文学法。

根据我們的多次实际計算經驗，对項数为10—20的較短观測系列，建議采用最小二乘法，对項数多于17—20—30項的較長系列，建議采用总和法。

本書是用来紀念水文工程师帕夫尔·阿列克桑德罗維奇·叶菲莫維奇的，他的許多見解与实际意見在本書中得到了反映，从設計莫斯科运河共同工作时期开始，本書作者在許多水文研究中曾利用过他的見解与意見。

讀者如認为需要对本書提出关于各种修改和补充的批評意見和願望，請寄莫斯科水閘河岸街10号(Москва, Шлюзовая набережная, 10.)国立动力出版社，作者將至为感謝。

作 者

目　　录

第一章　关于变差計算的一般概念
(按水文調查中的現代应用范圍)

野外水文調查及勘測就是要研究水文現象的若干基本組成部分，对于这些組成部分要用各种各样的水文測驗方法来进行观測。野外水文測驗資料的室內整編作業將在專門的手册、指令等中提到，因此，这里就不談这一問題。适于用变差計算法进行整編而使观測資料系統化的方法，我們从以下兩个方面来說明:

1)將一个分佈系列的研究的最終特征，列成保証率曲綫的形式;

2)將各相关系列的关系的研究的最終特征，列成相关系数和回归方程的形式。

1. 分佈系列

a)　分析分佈系列时的基本概念

經驗(統計或斯篤哈斯蒂)总体　对联合为統計总体的实际观測資料加以整編的工作，是变差計算法的基础。統計总体是数値(年平均逕流量或多年間的逕流量等)的一个变異系列，系列的每一項都是用数字来估計观測瞬間的現象。这种系列就叫作“分佈系列”，而系列中每一个單独項就叫作“变異項”。

統計总体分为“一般”(“全部”)总体和“部分”总体(或譯

为样本，或样品——譯註）。“一般”总体是由該系列一切可能的变異項組合而成，“部分”总体是指取某一“一般”总体的一部分而說的。实际上，多半不得不处理“部分”总体，因为几乎永远也不可能搜集到“一般”总体。

变異計算可以提供一些方法，使用这些方法，根据部分总体可以得到关于一般总体的資料，即接近实际的理論总体的資料。

理論总体 为了較全面地研究某一現象或为了証实所得到的規律的正确性起見，部分总体（实际观測点）永远必須伴随着相应的保証率曲綫（“一般”理論总体）。这种相应的保証率曲綫是根据經驗分佈系列的各計算参数（算术平均値、变差系数与偏差系数）来構成的。

統計变数，更确切些說是斯篤哈斯蒂变数 依各因素綜合来定的变数（此变数由于同一原因而具有帶一定或然性的不同数值，并發生变化），称为統計变数，更确切些說称为斯篤哈斯蒂变数，分佈系列就是由这类变数組成，水文学中大多数变数都屬于这一类。

概率 某一事件出現的情形的数目与其一切可能情形的数目的比率，就叫作这一事件的概率。例如，如果在125—130米^3/秒范圍內的年平均流量，在50年观測期間內（$n=50$）遇到7次，則在125—130米^3/秒范圍內的流量的重現概率为$\frac{7}{50}=0.14$或14%。

随着观測年数的增多，所研究的現象（在目前情形下是年平均流量）在一定間隔內的頻率，將接近概率。

6）整理經驗系列时的基本参数

为了选擇理論保証率曲綫起見，經驗分佈系列

$$x_1, x_2, \cdots\cdots, x_i \cdots\cdots\cdots, x_n$$

应由下列各参数表明。

系列 x 中各項的最大值为 $x_{макс}$，最小值为 $x_{мин}$。

系列各項的变幅(变化范圍)

$$A_x = x_{макс} - x_{мин}. \tag{1}$$

系列 x 的算术平均值:

$$\overline{x} = \frac{\Sigma x_n}{n_x}, \tag{2}$$

式中 n_x——系列 x 的項数。

标准差(离散度)σ_x表明了所研究的系列的各項对算术平均值的离差的平均数值大小。

显然，如果 $x_i > \overline{x}$，則所得到的个别离差值將具有正号(+)，如果 $x_i < \overline{x}$，則將具有負号(−)；离差的总和等于零(如果所进行的計算具有足够的精确度)，即

$$\Sigma \pm (x_i - \overline{x}) = [\Sigma + (x_i - \overline{x})] + [\Sigma - (x_i - \overline{x})] = 0$$

为了避免±符号对于离散度計算的影响起見，采用一种人为的方法：將分佈系列各項对算术平均值的每項离差，即所謂中心离差，暫时加以平方，这样結果就得到一个新的数字系列：

$$(x_1 - \overline{x})^2; (x_2 - \overline{x})^2 \cdots (x_i - \overline{x})^2 \cdots (x_n - \overline{x})^2 = \Sigma (x_i - \overline{x})^2.$$

这一个新系列將只由正的数值所組成。

为了求得平方离差的平均值起見，將所得到的总和除以系列的項数 n_x：

$$当 n_x > 30 时 \ \frac{\Sigma (x_i - \overline{x})^2}{n_x}$$

或　　　　　　当$n_x<30$时　$\frac{\Sigma(x_i-\bar{x})^2}{n_x-1}$

可以用符号X来代替中心离差，

$(x_1-\bar{x})=X_1$；　$(x_i-\bar{x})=X_i$等。

相应地$(x_1-\bar{x})^2=X_1^2$；　$(x_i-\bar{x})^2=X_i^2$等。

因此，在一切含有中心离差的公式中，都可以用符号X来代替中心离差。

为了得到分佈系列各項对算术平均值的平均离差(以一次方計)，需要取平方根，即將它改成标准差(离散度)的形式：

$$\sigma_x=\pm\sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n_x}}=\pm\sqrt{\frac{\Sigma X_i^2}{n_x}}, \tag{3}$$

或

$$\sigma_x=\pm\sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n_x-1}}=\pm\sqrt{\frac{\Sigma X_i^2}{n_x-1}}. \tag{3'}$$

通常，标准差的数值比簡單的平均离差要大一些，因为在个别項乘方时，較大离差增加得特别厉害。

"标准差"的数值，是以分佈系列为表示形式的、所研究的現象的变化性(变異性)的度量，并且是分隔較大离差与較小离差的界限；这样，数值处在$\pm\sigma_x$范圍內的系列各項对該系列来說是"正常"(典型)値，而其他各項都屬于極端値和需要專門研究的数值。

典型离差的界限可以解析地表示成下列形式

$$\bar{x}\pm\sigma_x\gtrless x_i. \tag{4}$$

算术平均值与标准差，都是有名数，即有因次的数值，标准差只是所研究的那个具体的分佈系列的各項离差的指标，并不能作出某种一般結論，例如，比較对某一現象在不

同观測时期所得到的最終数据，比較在不同河流或不同断面上所得到的結論等等。

为了消除这一項重要的缺点，数理統計学中采用变差（或变化性）系数。

变差系数是所研究的分佈系列的标准差与算术平均值的比，即

$$C_{v_x}=\frac{\sigma_x}{\bar{x}}=\frac{1}{\bar{x}}\sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n_x[\text{或}n_x-1]}}. \qquad (5)$$

变差系数是一个無因次的数值，它以一种抽象的表示法来表明一个具体的分佈系列的特征，使得有可能作各种比較的結論，用与其他类似研究相比較的方法来决定規律性等。

例如，假設系列 x 的变差系数 C_{v_x} 大于系列 x' 的 $C_{v_{x'}}$，而同时兩个系列都以同样的算术平均值为特征，即 $\bar{x}\simeq\bar{x}'$ 时，則这就意味着系列 x 的各項比系列 x' 的各項具有較大的变化。

分佈系列的偏差系数（偏斜度）C_{s_x}。以同样的或接近的算术平均值（即 $\bar{x}\simeq\bar{x}'$）和变差系数（$C_{v_x}\simeq C_{v_{x'}}$）为特征的兩个系列：x 和 x'，在系列中可能包括对于算术平均值的位置不同的若干个别項。这样的一些系列以偏差系数来区别。

假如正离差与負离差重現得同样頻繁，而分佈曲綫（第一章§3）（保証率曲綫是按分佈曲綫来繪制的）在中央具有最大的縱座标，而在均值兩旁等距离的縱座标彼此相等，則这种曲綫就叫作对称分佈曲綫（圖1，a与6）。对称分佈曲綫是最常遇到的一般形式的不对称分佈曲綫的一个特例。从頂点（峯部）算起，右部大于左部的分佈曲綫（在研究水文系列时最常遇到这种情形），具有正的偏斜度，而左部大于右部的分佈曲綫，具有負的偏斜度。

为了計算偏差系数起見，需要把各中心离差自乘三次方，这样就有可能查明正偏差与負偏差。

將这些离差的三次方总加起来时，可能有三种情形：

1. $\Sigma(x_i-\bar{x})^3=0$，对称系列，

2. $\Sigma(x_i-\bar{x})^3>0$，正偏系列，

3. $\Sigma(x_i-\bar{x})^3<0$，負偏系列。

所算得的总和是名数(有因次的数值)，为了决定其平均值并为了从有因次的偏差值轉变成叫作偏差系数 C_{s_x} 的無因次式起見，將上述总和除以乘积 $n_x\sigma_x^3$：

$$C_{s_x}=\frac{\Sigma(x_i-\bar{x})^3}{n_x\sigma_x^3}=\frac{\Sigma X_i^3}{n_x\sigma_x^3}. \tag{6}$$

在研究分佈系列时，利用偏差系数可以查明各种規律、各种对照手續和其他分析与結論。

应該指出，当系列項数少于 100 时，偏差系数的計算不能得到可靠的結果；某些著者認为在实际計算中，可以使用較短的系列(25 至 50 年)为可靠与否的限度[参考文献 13]。

在研究逕流量和流量等的水文計算中，通常不計算偏差系数，而是有条件的設它等于变差系数乘 2：

$$C_{s_x}=2C_{v_x}. \tag{6'}$$

对于降雨流量[参考文献20]：

$$C_s=4C_v. \tag{6''}$$

按[参考文献20]，以下列不等式来决定 C_s 的極限值：

$$\frac{2C_{v_x}}{1-\frac{x_{мин}}{\bar{x}}}\geqslant C_{s_x}\geqslant 2C_{v_x}. \tag{6'''}$$

对于某一个具体的、局部分佈系列(例如，年逕流量、

流量等数值的系列），以按照上述各公式所算得的 $\bar{x},\sigma_x,C_{v_x}$ 与 C_{s_x} 的数值为根据，决不能对一般現象的特征作出决定性的結論。必須算出这些参数的“統計誤差”，并把“統計誤差”加在上述各参数的数値中作为修正数。

統計誤差 数理統計学已拟制出一些决定上述各参数計算精度界限的方法，用术語来講，計算精度界限就叫作“統計誤差”。

应該着重指出，統計誤差無論如何决不反映“由于进行水文測驗工作本身的不精确”所引起的实际誤差，这一点在任何情况下都不能混淆起来（“誤差”这个詞在这里只是变差計算中的一个术語）。

統計誤差理論給出一种指示，即按照一个具体的分佈系列的参数 $\bar{x},\sigma_x,C_{v_x},C_{s_d}$ 的数值，如何决定这些参数在一般总体中的数值界限，也就是給出这些参数的最終的（理論的）变幅。

概率理論确定，所研究的分佈系列的各項变化性（离散度）越小，而系列的項数越多，則統計誤差的数値就越小，也就是說，从局部分佈系列中所求到的参数值就越接近这一参数在一般全部总体中的数値。

算术平均值的平均誤差：

$$E_x=\pm\frac{\sigma_x}{\sqrt{n_x}}. \tag{7}$$

数值 E_x，表示在同类观測資料中但分組不同时所算出的若干个算术平均値的可能变化的平均变幅，它是依标准差 σ_x 大小与观測次数多寡而定的精度决定。

变化的最大限度，应該增大到平均誤差乘三（$3E_x$）。

标准差的平均誤差：

$$E_{\sigma_x}=\pm\frac{\sigma_x}{\sqrt{2n_x}} \tag{8}$$

这种誤差表示在完全一致的原始資料中，在不同的数值下多次重复計算时所得到的离散度可能变化的某一平均变幅。

例如，計算求到$E_{\sigma_x}=\pm0.15$，可以証明在其他类似研究(这些研究实际上并未进行)中，离散度σ_x的数值可能是与这个数值差±0.15的数值。

标准差的誤差小于算术平均值的誤差，因此，在一切条件下，决定的σ_x比决定的$\bar{x}$有較大的精度。

变差系数的平均誤差：

$$E_{v_x}=\pm\frac{C_{v_x}}{\sqrt{2n_x}}\sqrt{1+2C_{v_x}^2}. \tag{9}$$

偏差系数的平均誤差(当$n>100$时)：

$$E_{s_x}=\pm\sqrt{\frac{6}{n_x}}. \tag{10}$$

C.H.克里茨基与M.Φ.明克尔随后分析了分佈曲綫，在$C_s=2C_v$的条件下导出下列求标准誤差的公式：

$$E_{\sigma_x}=\pm\frac{\sigma_x}{\sqrt{2n_x}}\sqrt{1+3C_{v_x}^2}, \tag{8'}$$

$$E_{s_x}=\pm\sqrt{\frac{6}{n_x}(1+6C_{v_x}^2+5C_{v_x}^3)}. \tag{10'}$$

用計算求到局部分佈系列的誤差之后，可以假設，一般全部总体的参数的变化界限在下列范圍內：

$$\bar{x}\pm E_x;\ \sigma_x\pm E_{\sigma_x};\ C_{v_x}\pm E_{v_x};\ C_{s_x}\pm E_{s_x}.$$

在这些数值以上及以下的数值就是这些参数的極端值。

或然誤差 上述計算平均“統計誤差”的各公式，是对于

数理統計学中著名的叫作“按高斯法則的正态分佈情形”所推演出来的，并不完全适用于河川逕流变化性的研究，对于河川逕流来說，不对称的分佈曲綫是具有代表性的(圖I，a与$б$)，因为所有情形的一半中(50%)，其偏差都小于 $\pm E_{50\%}$，因此，应該把統計誤差公式看成是計算“或然誤差”以作为修正数的近似公式。

A.K.米特罗波尔斯基〔参考文献11，第295頁〕教授認为，从理論和实踐观点出發，都說明用基本誤差来作为可靠性的度量，比用或然誤差有很多的优点。

为了求到或然誤差($E_{50\%}$)的公式起見，要將平均基本誤差乘上0.67449〔参考文献9〕。

这样，算术平均值的或然誤差为：

$$E_{x_{50\%}}=\pm 0.674\frac{\sigma_x}{\sqrt{n_x}};\tag{7'}$$

标准差的或然誤差

$$E_{\sigma_{50\%}}=\pm 0.674\frac{\sigma_x}{\sqrt{2n_x}};\tag{8''}$$

变差系数的或然誤差

$$E_{v_{50\%}}=\pm 0.674\frac{C_{v_x}}{\sqrt{2n_x}}\sqrt{1+2C_{v_x}^2}.\tag{9'}$$

附註： 在水文統計計算中，关于算术平均值（逕流正常值$\bar{q}$）的或然誤差問題，C.H.克里茨基与M.Ф.明克尔〔参考文献9〕認为算术平均值应該在$\bar{q}\pm 4E_{50\%}$的范圍內，但是更可能的是$\bar{q}$在較狹窄的范圍內：$q\pm E_{50\%}$。

峯度指标 实际的分佈曲綫(第一章§3)可以用偏差指标与峯度指标来表示其特征。这兩項指标可以决定某一分佈曲綫与正态分佈曲綫相比較的程度、特性和形狀。

关于偏差的概念(圖 1, a与 $б$)并不能概括破坏正态分佈的全部典型情形。例如，分佈多边形的峯点显著地突出正态曲綫的范圍之外(圖1, $в$)，大多数变異項都集中在系列的中部，因此，某些个别变異項就散在兩旁而距算术平均值有較大的距离；或是观察到中央部分低落(圖1, г)，因此，大多数变異項就在距算术平均值的兩側均匀分佈。系列保持完全对称时，对于正态分佈的破坏的性質就叫作峯度。

分佈多边形高聳时所表示的峯度(圖1, $в$)叫作正峯，而系列中部凹陷或分佈多边形峯部低而平时叫作負峯(圖1, г)。系列各項分佈的差異表現在峯度公式所包括的中心离差四次方总和$\Sigma(x_i-\bar{x})^4=\Sigma X_i^4$的数值上。对正态分佈曲綫加以数学分析，确定了表明峯度特性与程度的極限值等于 3：

$$\frac{\Sigma(x_i-\bar{x})^4}{n\sigma_x^4}=\frac{\Sigma X_i^4}{n\sigma_x^4}=3.$$

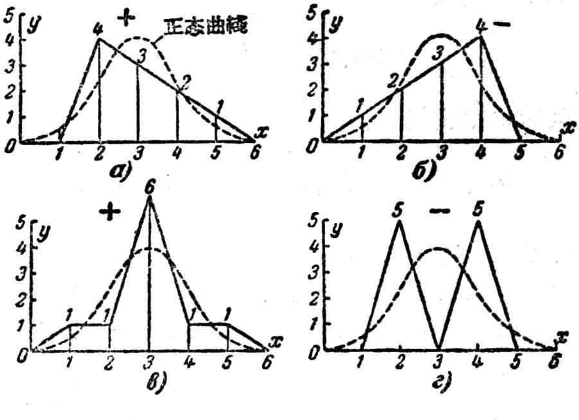

圖 1 实际分佈曲綫与正态曲綫的比較

a—正偏分佈曲綫；$б$—負偏分佈曲綫；$в$—正峯分佈曲綫；$г$—負峯分佈曲綫。

这样，峯度指标ε_{s_x}就按下列公式計算

$$\varepsilon_{s_x}=\frac{\Sigma(x_i-\bar{x})^4}{n\sigma_x^4}-3=\frac{\Sigma X_i^4}{n\sigma_x^4}-3. \qquad (11)$$

正峯时所得峯度指标的符号为(+)(圖 1, в)，負峯时所得峯度指标的符号为(−)(圖 1, г)。

按照以上所述，偏差系数[公式(6)]与峯度系数[公式(11)]是实际分佈曲綫的形狀与特性的基本指标。

*　　*　　*

对于一个系列的各項用解析方法来进行研究，即算出上述各种参数之后，可以：

1)根据观測的局部系列的資料，得到所研究的現象（例如，河川年逕流量等）的一般理論特征，并把它看成是反映整个該現象的一般全部規律的計算特征；

2)按照上述各参数的数值：

а)繪制理論的保証率曲綫；

б)比較資料，例如对于不同的河流的資料进行比較；

в)表明同一河流在不同时期的資料特征，并加以比較；

г)將具有多年观測的河流特征換算为观測年数較少的另一条河流的特征等。

在具有長期观測($n\geqslant 17—20—30$年)时，以及在具有短期观測($n\leqslant 10—20$年)时，甚至在完全沒有观測时，都可以进行参数的計算及其可靠程度的估算，在沒有观測时就要利用經驗公式。为了計算正常值所需要的年数应符合于以百分数計的規定誤差。

*　　*　　*

对于多年逕流特征来說，观測系列越長，則誤差就越少，因为其算术平均值$\bar{x}$就越接近眞值的原故。

算术平均值的平均誤差，以算术平均值 $\bar{x}$ 的百分数計［参考文献 19，第 99 頁］：

$$E_x\% = \pm \frac{100\frac{\sigma_x}{\bar{x}}}{\sqrt{n_x}} = \pm 100\frac{C_{v_x}}{\sqrt{n_x}} \tag{7''}$$

由此

$$n_x = \frac{10^4 C_{v_x}^2}{(E_x\%)^2}. \tag{12}$$

Б. Д. 查依科夫和 С. Ю. 别林科夫［参考文献 19］按照这一公式編制出一張表，已知規定的平均誤差 $E_x\%$，并算出数值 C_{v_x} 之后，就可以确定对于規定誤差所需要的年数。

表 1

依算术平均值的平均誤差与变差系数而定的观测年数

C_{v_x}	E_x，%							
	±4	±5	±6	±7	±8	±9	±10	±20
0.15	14	9	6	5	4	3	2	1
0.20	25	16	11	8	6	5	4	1
0.25	39	25	17	13	10	8	6	2
0.30	56	36	25	19	14	11	9	2
0.35	76	49	33	25	19	15	12	3
0.40	100	64	44	33	25	20	16	4
0.45	126	81	55	42	32	25	20	5
0.50	156	100	69	50	39	31	25	6
0.55	189	121	83	62	47	38	30	8
0.60	225	144	99	74	56	45	35	9

当数值 n>10—15 年时，可以直接計算变差系数 C_v；但是，假如規定变差系数的誤差与算术平均值的誤差同样大小时，也可以用另一种方法来求到所必須的观测年数。