Извѣстія Императорской Академіи Наукъ. — 1913.
(Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg).

Примѣръ статистическаго изслѣдованія надъ текстомъ „Евгенія Онѣгина“ иллюстрирующій связь испытаній въ цѣпь.

А. А. Марковъ.

(Доложено въ засѣданіи Физико-Математическаго Отдѣленія 23 января 1913 г.).

Наше изслѣдованіе относится къ послѣдовательности 20 000 русскихъ буквъ, не считая ъ и ь, въ романѣ А. С. Пушкина «Евгеній Онѣгинъ», которая заполняетъ всю первую главу и шестьнадцать строфъ второй.

Эта послѣдовательность доставляетъ намъ 20 000 связанныхъ испытаній, каждое изъ которыхъ даетъ гласную или согласную букву.

Соотвѣтственно этому мы допускаемъ существованіе неизвѣстной постоянной вѣроятности p буквѣ быть гласной и приближенную величину числа p ищемъ изъ наблюденій, считая число появившихся гласныхъ и согласныхъ буквъ. Кромѣ числа p мы найдемъ, также изъ наблюденій, приближенныя величины двухъ чиселъ p_1 и p_0 и четырехъ чиселъ $p_{1,1}$, $p_{1,0}$, $p_{0,1}$, $p_{0,0}$, представляющихъ такія вѣроятности: p_1 — гласной слѣдовать за гласной, p_0 — гласной слѣдовать за согласной, $p_{1.1}$ — гласной слѣдовать за двумя гласными, $p_{1,0}$ — гласной слѣдовать за согласной, которой предшествуетъ гласная, $p_{0,1}$ — гласной слѣдовать за гласной, которой предшествуетъ согласная и, наконецъ, $p_{0,0}$ — гласной слѣдовать за двумя согласными.

Эти обозначенія согласованы съ принятыми въ статьѣ моей «Объ одномъ случаѣ испытаній связанныхъ въ сложную цѣпь»; при ссылкѣ же на статью «Изслѣдованіе замѣчательнаго случая зависимыхъ испытаній» надо p_0 приравнять p_2. Противоположныя вѣроятности, буквѣ быть согласной, обозначимъ, какъ принято нами, буквою q съ тѣми же значками.

Разыскивая число p, мы находимъ для него сначала 200 приближенныхъ величинъ, изъ которыхъ затѣмъ выводимъ среднюю ариѳметическую.

А именно, мы разбиваемъ всю послѣдовательность 20 000 буквъ на 200 отдѣльныхъ послѣдовательностей по 100 буквъ и считаемъ, сколько гласныхъ въ каждой сотнѣ буквъ: мы получаемъ 200 чиселъ, которыя, по раздѣленіи на 100, даютъ 200 приближенныхъ величинъ p.

При счетѣ числа гласныхъ мы имѣемъ въ виду сохранить возможность образовать другія соединенія по 100 буквъ; каждую изъ нашихъ сотенъ мы располагаемъ въ квадратъ по десяти строкъ и десяти столбцовъ, сохраняя порядокъ буквъ:

1	2	3	4	5	6	7	8	9	10
11,	12,	13,	14,	15,	16,	17,	18,	19,	20
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
91,	92,	93,	94,	95,	96,	97,	98,	99,	100.

Считаемъ сколько гласныхъ въ каждомъ столбцѣ, въ отдѣльности, и соединяемъ числа по два:

$$1^{\text{ое}} \text{ и } 6^{\text{ое}},\quad 2^{\text{ое}} \text{ и } 7^{\text{ое}},\quad 3^{\text{ье}} \text{ и } 8^{\text{ое}},\quad 4^{\text{ое}} \text{ и } 9^{\text{ое}},\ 5^{\text{ое}} \text{ и } 10^{\text{ое}}.$$

Мы получаемъ такимъ образомъ для каждой сотни буквъ пять чиселъ, обозначаемыхъ нами символами

$$(1,6),\quad (2,7),\quad (3,8),\quad (4,9),\quad (5,10);$$

сумма ихъ

$$(1,6) + (2,7) + (3,8) + (4,9) + (5,10)$$

равна числу гласныхъ этой сотни.

Соединяя же по 500 буквъ вмѣстѣ, мы можемъ образовать новыя пять сотенъ буквъ: первую — изъ первыхъ и шестыхъ столбцовъ, вторую — изъ вторыхъ и седьмыхъ столбцовъ и т. д.

Число гласныхъ въ этихъ новыхъ сотняхъ опредѣляется, очевидно, суммами

$$\Sigma\,(1,6),\quad \Sigma\,(2,7),\quad \Sigma\,(3,8),\quad \Sigma\,(4,9),\quad \Sigma\,(5,10),$$

состоящими изъ соотвѣтствующихъ пяти слагаемыхъ.

Результаты нашего счета приведены въ сорока табличкахъ, каждая изъ которыхъ содержитъ: въ первой строкѣ — пять чиселъ (1,6) и ихъ сумму, во второй строкѣ — пять чиселъ (2,7) и ихъ сумму и т. д., а въ послѣдней строкѣ — число гласныхъ въ первой сотнѣ, во второй сотнѣ и т. д. и наконецъ число гласныхъ во всѣхъ пяти сотняхъ, уменьшенное для сбереженія мѣста на 200.

6 8 11 11 13 49	16 11 9 8 7 51	14 12 7 3 6 42	5 11 10 6 10 42	10 6 6 6 7 35
12 11 7 7 5 42	4 8 9 11 10 42	5 5 11 9 11 41	12 8 8 11 7 46	9 12 15 6 9 51
6 6 6 7 13 38	9 9 9 7 10 44	8 10 6 10 7 41	7 7 12 10 9 45	9 3 6 10 9 37
8 10 11 9 4 42	12 9 6 10 7 44	11 11 8 3 10 43	8 12 7 9 9 45	9 11 8 5 6 39
10 11 5 10 8 44	3 8 10 8 9 38	4 4 11 14 8 41	12 8 10 9 8 47	9 10 10 10 9 48
42 46 40 44 43 15	**44 45 43 44 43** 19	**42 42 43 39 42** 8	**44 46 47 45 43** 25	**46 42 45 37 40** 10
8 7 8 7 10 40	11 11 8 7 7 44	11 10 10 12 6 49	12 9 8 10 10 49	8 9 9 5 8 39
10 9 9 8 8 44	9 6 10 11 11 47	4 4 9 7 9 33	3 10 12 9 10 44	7 9 9 11 7 43
8 9 8 8 8 41	12 9 9 5 6 41	11 13 6 9 10 49	11 11 6 11 10 49	10 6 6 9 9 40
10 6 13 6 12 47	10 8 6 11 11 46	6 7 11 8 6 38	10 8 11 6 7 42	7 8 15 6 9 45
8 12 5 13 6 44	7 6 8 9 8 38	8 6 10 7 12 43	6 8 7 9 6 36	11 7 6 11 10 45
44 43 43 42 44 16	**49 40 41 43 43** 16	**40 40 46 43 43** 12	**42 46 44 45 43** 20	**43 39 45 42 43** 12
7 7 7 7 9 37	12 7 7 6 8 40	7 4 11 5 7 34	5 5 7 5 9 31	8 6 5 14 11 44
9 13 6 8 4 40	6 8 7 10 8 39	11 14 9 11 9 54	12 6 10 10 8 46	8 12 10 7 4 41
9 7 11 12 14 53	9 10 10 8 7 44	7 6 9 8 9 39	8 14 11 11 10 54	8 10 9 8 14 49
7 11 8 9 7 42	9 5 6 7 7 34	10 9 8 10 5 42	4 3 9 5 9 30	9 5 9 9 6 38
8 10 10 11 9 48	7 11 9 13 7 47	11 10 8 9 11 49	13 14 9 11 7 54	8 13 11 5 10 47
40 48 42 47 43 20	**43 41 39 44 37** 4	**46 43 45 43 41** 18	**42 42 46 42 43** 15	**41 46 44 43 45** 19
10 9 13 6 12 50	4 11 10 12 5 42	5 11 10 6 5 37	4 4 10 11 5 34	13 11 13 10 10 57
8 8 8 9 5 38	14 9 8 7 14 52	8 9 8 10 10 45	6 12 9 8 10 45	7 10 9 6 2 34
10 10 8 9 10 47	4 8 9 8 4 33	8 8 6 9 9 40	13 4 10 8 6 41	8 8 7 8 12 43
7 9 10 7 10 43	8 14 11 12 6 51	10 6 9 7 6 38	7 10 7 12 11 47	9 11 9 10 6 45
9 8 3 11 7 38	11 6 7 4 14 42	11 9 8 10 12 50	9 13 8 1 8 39	6 3 7 9 9 34
44 44 42 42 44 16	**41 48 45 43 43** 20	**42 43 41 42 42** 10	**39 43 44 40 40** 6	**43 43 45 43 39** 13
11 6 8 9 5 39	10 10 4 7 9 40	10 8 7 8 8 41	10 3 11 13 5 42	8 8 13 5 8 42
6 10 6 8 13 43	11 10 13 13 9 56	6 9 9 8 7 39	7 11 9 7 10 44	9 10 7 14 9 49
10 5 11 11 6 43	10 7 5 9 6 37	15 9 11 13 9 57	10 10 4 7 7 38	9 11 6 8 7 41
9 12 6 8 10 45	10 5 8 10 10 43	5 10 5 4 7 31	7 7 14 13 7 48	7 9 12 6 9 43
7 11 9 10 10 47	6 13 10 5 6 40	8 9 10 12 9 48	11 9 9 6 15 50	10 9 9 12 9 49
43 44 40 46 44 17	**47 45 40 44 40** 16	**44 45 42 45 40** 16	**45 40 47 46 44** 22	**43 47 47 45 42** 24
12 7 12 5 12 48	10 14 7 6 6 43	9 6 7 10 5 37	12 13 5 9 11 50	5 11 8 12 10 46
10 8 5 13 4 40	4 6 8 10 14 42	11 10 7 8 9 45	7 7 10 5 8 37	12 8 9 8 6 43
10 13 8 7 9 47	13 6 12 8 5 44	10 10 9 9 10 48	7 7 9 14 7 44	8 11 9 8 7 43
9 4 12 6 9 40	7 13 5 8 10 43	8 6 12 10 10 46	12 13 7 8 10 50	8 5 7 11 8 39
4 12 9 9 8 42	8 5 15 10 9 47	9 11 8 5 11 44	4 4 12 11 9 40	11 11 10 6 8 46
45 44 46 40 42 17	**42 44 47 42 44** 19	**47 43 43 42 45** 20	**42 44 43 47 45** 21	**44 46 43 45 39** 17
9 11 10 6 13 49	5 9 7 10 6 37	8 6 8 7 14 43	7 9 8 6 7 37	9 11 11 8 8 47
9 8 6 8 6 37	10 9 11 7 7 44	8 14 13 8 4 47	9 8 6 10 11 44	10 8 5 9 10 42
7 7 12 10 9 45	11 11 11 10 8 51	12 4 6 9 11 42	10 9 10 8 10 47	6 8 16 12 11 53
12 12 6 8 8 46	7 7 5 10 10 39	6 8 9 10 8 41	8 7 4 9 4 32	12 11 5 7 8 43
5 7 9 11 4 36	13 8 9 8 10 48	6 8 11 8 6 39	11 8 10 8 9 46	6 5 9 10 8 38
42 45 43 43 40 13	**46 44 43 45 41** 19	**40 40 47 42 43** 12	**45 41 38 41 41** 6	**43 43 46 46 45** 23
5 7 4 3 7 26	4 7 9 11 10 41	10 8 7 8 7 40	12 10 11 4 5 42	12 13 6 6 10 47
14 10 13 9 5 51	10 7 9 4 9 39	10 8 11 10 7 46	5 9 10 11 11 46	6 3 10 10 4 33
7 8 6 8 9 38	8 13 9 12 10 52	6 11 11 10 10 48	10 8 10 7 13 48	11 11 9 7 14 52
7 10 9 5 9 40	7 5 7 7 12 38	12 8 7 6 5 38	11 8 8 11 5 43	5 8 8 9 9 39
9 10 11 16 7 53	13 10 10 9 5 47	5 9 11 12 11 48	4 8 8 9 11 40	11 6 11 12 7 47
42 45 43 41 37 8	**42 42 44 43 46** 17	**43 44 47 46 40** 20	**42 43 47 42 45** 19	**45 41 44 44 44** 18

Остановимся на совокупности чиселъ

42, 46, 40, 44, 43, 44, 45, 43, . . .

стоящихъ въ послѣднихъ строкахъ нашихъ 40 табличекъ и показывающихъ, сколько находится гласныхъ въ послѣдовательныхъ сотняхъ текста:

1) мой дядя самых честных правил когда не в шутку занемог он уважат себя заставил и лучше выдумат не мог его примѣр другим на (42 гласныхъ)

2) ука но боже мой какая скука с болным сидѣт и ден и ноч не отходя ни шагу проч какое низкое коварство полуживаго забавлят ем (46 гласныхъ)

и т. д.

Считая, сколько разъ въ этой совокупности встрѣчается каждое число составляемъ новую небольшую таблицу

37	38	39	40	41	42	43	44	45	46	47	48	49
3	1	6	18	12	31	43	29	25	17	12	2	1

Здѣсь въ первой строкѣ приведены всѣ числа, входящія въ нашу совокупность, а подъ ними, во второй строкѣ, указано, сколько разъ они встрѣчаются.

При помощи этой таблицы легко находимъ ихъ среднее арифметическое

$$43 + \frac{29+25\times2+17\times3+12\times4+2\times5+6-31-12\times2-18\times3-6\times4-5-3\times6}{200} = 43,19$$

и отсюда выводимъ

$$p \doteq 0,4319 \doteq 0,432.$$

Вычисляемъ сумму квадратовъ ихъ отклоненій отъ 43,2; она оказывается равною

1022,8,

что по раздѣленіи на 200 даетъ намъ число

5,114.

которое можно принять за приближенную величину математическаго ожиданія квадрата отклоненія любого изъ нашихъ 200 чиселъ отъ ихъ общаго математическаго ожиданія, приблизительно равнаго 43,2. Наконецъ число

$$\frac{5,114}{200} = 0,02557$$

представляетъ приближенную величину математическаго ожиданія квадрата погрѣшности въ опредѣленіи $100\,p$ равенствомъ

$$100\,p \doteq 43,2.$$

Такое заключеніе соединено съ обычнымъ предположеніемъ способа наименьшихъ квадратовъ, что мы имѣемъ дѣло съ независимыми величинами. Это предположеніе, въ данномъ случаѣ, оправдывается не хуже, чѣмъ во многихъ другихъ, ибо связь между числами, по способу ихъ полученія, весьма слаба.

Можно подмѣтить также нѣкоторую согласованность нашихъ результатовъ съ извѣстнымъ закономъ погрѣшности, связаннымъ съ именами Гаусса и Лапласа; напримѣръ, величина называемая вѣроятною погрѣшностью у насъ приблизительно равна

$$0{,}67 \, . \sqrt{5{,}11} \doteq 1{,}5$$

и соотвѣтственно этому между

$$43{,}2 - 1{,}5 = 41{,}7 \quad \text{и} \quad 43{,}2 + 1{,}5 = 44{,}7$$

находится 103 числа, т. е. около половины ихъ: 31 разъ число 42, 43 раза число 43 и 29 разъ число 44.

Независимости нашихъ величинъ соотвѣтствуетъ тотъ фактъ, что, соединяя ихъ по двѣ, по четыре и по пяти и вычисляя для этихъ 100, 50 и 40 комбинацій суммы квадратовъ ихъ отклоненій отъ

$$86{,}4, \quad 172{,}8 \quad \text{и} \quad 216,$$

мы получаемъ числа

$$827{,}6 \quad 975{,}2, \quad 1004,$$

которыя не очень сильно отличаются отъ ранѣе найденнаго числа

$$1022{,}8.$$

Переходя отъ сотенъ испытаній къ отдѣльнымъ испытаніямъ, замѣчаемъ, что число

$$\frac{5{,}114}{100} = 0{,}05114$$

сильно отличается отъ

$$0{,}432 \times 0{,}568 = 0{,}245376:$$

коэффиціентъ дисперсіи (мы не много отступаемъ отъ обычнаго словоупотребленія, согласно которому слѣдовало бы извлечь квадратный корень изъ числа, названнаго нами коэффиціентомъ дисперсіи) оказывается равнымъ

$$\frac{5114}{24537{,}6} \doteq 0{,}208,$$

т. е. составляетъ около $\frac{1}{5}$, что прекрасно объясняется связанностью нашихъ испытаній.

Для выясненія этой связи, хотя бы и не полнаго, намъ можетъ послужить приближенное вычисленіе вышеупомянутыхъ вѣроятностей p_1 и p_0.

Просматривая весь текстъ изъ 20 000 буквъ, мы считаемъ, сколько въ немъ встрѣчается послѣдовательностей

гласная, гласная;

получаемъ число 1104, которое по раздѣленіи на число всѣхъ гласныхъ въ текстѣ даетъ для p_1 такую приближенную величину

$$\frac{1104}{8638} \doteqdot 0{,}128.$$

Подобнымъ же образомъ, считая число послѣдовательностей

согласная, согласная

и дѣля его на 11362, мы могли бы найти приближенное значеніе q_0 и затѣмъ $p_0 = 1 - q_0$. Но можно замѣнить утомительный прямой счетъ слѣдующимъ. Вычитая 1104 изъ 8638 находимъ число согласныхъ

7534,

слѣдующихъ за гласными, а такъ какъ, кромѣ первой, всѣ согласныя должны слѣдовать за гласной или за согласной, то число послѣдовательностей

согласная, согласная

опредѣляется разностью

$$11361 - 7534 = 3827.$$

Отсюда тотчасъ получаемъ для p_0 такую приближенную величину

$$\frac{7534}{11361} \doteqdot \frac{7534}{11362} \doteqdot 0{,}663.$$

Мы видимъ, что вѣроятность буквѣ быть гласной значительно измѣняется, въ зависимости отъ того, предшествуетъ ей гласная или согласная, разность $p_1 - p_0$, обозначаемая нами буквою δ, оказывается равною

$$0{,}128 - 0{,}663 = -0{,}535.$$

Если мы допустимъ теперь, что наша послѣдовательность 20 000 буквъ образуетъ простую цѣпь, то при

$$\delta = -0{,}535$$

за теоретическій коэффиціентъ дисперсіи можно принять, согласно «Изслѣдованію замѣчательнаго случая зависимыхъ испытаній», число

$$\frac{1+\delta}{1-\delta} = \frac{465}{1535} \doteqdot 0{,}3;$$

конечно, это число не вполнѣ совпадаетъ съ полученнымъ нами раньше

$$0.208,$$

но, во всякомъ случаѣ, подходитъ къ нему ближе, чѣмъ число единица, соотвѣтствующее случаю независимыхъ испытаній.

Если же разсматривать нашу послѣдовательность какъ сложную цѣпь и примѣнить сюда выводы изслѣдованія «Объ одномъ случаѣ испытаній связанныхъ въ сложную цѣпь», то можно еще лучше согласовать теоретическій коэффиціентъ дисперсіи съ опытнымъ.

Для этого считаемъ, сколько въ нашей послѣдовательности находится комбинацій

гласная, гласная, гласная,

и

согласная, согласная, согласная;

число первыхъ комбинацій, по моему счету, оказывается равнымъ 115, а вторыхъ — 505. Дѣля эти числа на найденныя ранѣе

1104 и 3827,

получаемъ приближенныя равенства

$$p_{1,1} \doteq \frac{115}{1104} \doteq 0.104, \quad q_{0,0} \doteq \frac{505}{3827} \doteq 0.132.$$

Чтобы примѣнить теперь къ нашему случаю выводы только что упомянутой статьи, полагаемъ

$$p \doteq 0.432, \ q \doteq 0.568, \ p_1 = 0.128, \ q_1 = 0.872, \ p_0 = 0.663, \ q_0 = 0.337,$$
$$p_{1,1} = 0.104, \quad q_{0,0} = 0.132$$

и по этимъ числамъ находимъ

$$\delta = -0.535, \quad \varepsilon = \frac{-24}{872} \doteq -0.027, \quad \eta = -\frac{205}{663} \doteq -0.309.$$

Затѣмъ обращаемся къ выраженію коэффиціента дисперсіи

$$\frac{\{q(1-3\varepsilon)(1-\eta)+p(1-3\eta)(1-\varepsilon)-2(1-\varepsilon)(1-\eta)\}(1-\delta)+2(1-\varepsilon\eta)}{(1-\delta)(1-\varepsilon)(1-\eta)} =$$

$$= \frac{1+\delta}{1-\delta}\left\{\frac{1+\varepsilon}{2(1-\varepsilon)}+\frac{1+\eta}{2(1-\eta)}\right\}+\frac{(q-p)(\eta-\varepsilon)}{(1-\varepsilon)(1-\eta)},$$

которое соотвѣтствуетъ условіямъ той статьи и въ ней выведено.

Подставивъ сюда найденныя нами значенія

$$p, \ q, \ \delta, \ \varepsilon, \ \eta$$

и произведя выкладки, получаемъ для коэффиціента дисперсіи число

$$0{,}195,$$

которое настолько согласуется съ найденнымъ по общимъ правиламъ, независимо отъ нашихъ особыхъ предположеній, числомъ

$$0.208,$$

что большаго согласія едва ли можно требовать.

Нельзя, конечно, утверждать, что нашъ примѣръ удовлетворяетъ теоретическимъ условіямъ во всей полнотѣ; но, съ другой стороны, едва ли можно сомнѣваться, что отмѣченное нами согласіе чиселъ не случайно и связано съ извѣстною согласованностью теоретическихъ предположеній съ условіями примѣра.

Переходимъ къ другому, произведенному нами, распредѣленію 20 000 буквъ на сотни. Составляемъ для него таблицу повторяемости различныхъ чиселъ, подобную прежней

26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41
1	0	0	0	1	2	1	3	5	1	2	9	13	12	13	11

42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57
17	16	15	10	10	16	10	10	5	5	3	3	3	0	1	2

Среднее арифметическое изъ этихъ новыхъ 200 чиселъ равно прежнему

$$43.19.$$

Сумма же квадратовъ ихъ отклоненій отъ 43.2 значительно больше прежней; а именно, она равна

$$5788.8.$$

Здѣсь слѣдуетъ остановиться на условіи независимости величинъ, обычно соединяемымъ со способомъ наименьшихъ квадратовъ (см. главу VII моей книги «Исчисленіе вѣроятностей»); вспомнимъ, для чего нужно это условіе. Оно является необходимымъ при разысканіи вѣса окончательнаго результата выражаемаго равенствомъ (21) и при вычисленіи математическаго ожиданія W, которое даетъ намъ приближенную величину k

(см. мою книгу). Но это условіе окажется лишнимъ, если мы, во первыхъ, оставимъ въ сторонѣ вопросъ о вѣсѣ равенства (21) и, во вторыхъ, замѣнимъ ξ въ выраженіи W числомъ a, которое потомъ будемъ считать равнымъ a_0, пренебрегая разностью $a - a_0$. Тогда въ основу нашихъ сужденій лягутъ два равенства

$$\text{м. о. } \frac{p' x' + p'' x'' + \ldots + p^{(n)} x^{(n)}}{p' + p'' + \ldots + p^{(n)}} = a$$

и

$$\text{м. о. } \frac{p' (x' - a)^2 + p'' (x'' - a)^2 + \ldots + p^{(n)} (x^{(n)} - a)^2}{n} = k,$$

не требующія независимости величинъ

$$x', \; x'', \; \ldots, \; x^{(n)}.$$

На основаніи такихъ равенствъ, опираясь на законъ большихъ чиселъ, мы полагаемъ

$$a \doteqdot \frac{p' a' + p'' a'' + \ldots + p^{(n)} a^{(n)}}{p' + p'' + \ldots + p^{(n)}} = a_0$$

и

$$k \doteqdot \frac{\Sigma\, p^{(i)} (a^{(i)} - a)^2}{n} \doteqdot \frac{\Sigma\, p^{(i)} (a^{(i)} - a_0)^2}{n}.$$

Отпадаетъ только теорема о вѣсѣ окончательнаго результата, выражаемая извѣстнымъ равенствомъ (22): вѣсъ результата равенъ суммѣ вѣсовъ составляющихъ.

Въ данномъ случаѣ каждое изъ нашихъ 200 чиселъ представляетъ сумму почти независимыхъ величинъ; но зато сами суммы связаны но пяти, такъ что только сорокъ изъ нихъ можно считать независимыми. Мы имѣемъ 40 группъ по 500 буквъ; въ каждой сотнѣ нѣтъ смежныхъ буквъ текста, чѣмъ обусловливается отмѣченная нами независимость слагаемыхъ; зато въ каждой группѣ смежны буквы первой сотни съ буквами второй сотни, буквы второй сотни съ буквами первой и третьей и т. д., въ силу чего наши числа связаны по пяти, какъ сказано выше.

При такихъ условіяхъ, согласно приведеннымъ объясненіямъ, число

$$\frac{5788{,}8}{200} = 28{,}944$$

можно разсматривать какъ приближенную величину математическаго ожиданія квадрата отклоненія нашихъ новыхъ 200 чиселъ.

$$49, \; 42, \; 38, \; 42, \; 44, \ldots$$

отъ ихъ математическаго ожиданія, приблизительно равнаго

$$43{,}2.$$

И переходя отъ сотенъ буквъ (испытаній) къ отдѣльнымъ буквамъ, мы замѣчаемъ теперь, что число

$$0{,}28944$$

не очень сильно отличается отъ

$$0{,}432 \times 0{,}568 = 0{,}245376:$$

коэффиціентъ дисперсіи оказывается равнымъ

$$\frac{28944}{24537{,}6} \doteq 1{,}18.$$

Если же мы обратимся къ окончательному результату

$$43{,}19,$$

то математическое ожиданіе квадрата его погрѣшности нельзя уже выражать числомъ

$$\frac{28{,}944}{200} = 0.14472,$$

въ виду связи нашихъ чиселъ

$$49,\ 42,\ 38,\ 42,\ 44\ldots;$$

напротивъ это математическое ожиданіе можно, согласно результатамъ первоначальнаго распредѣленія буквъ на сотни, выразить, конечно приближенно, числомъ

$$\frac{5{,}114}{200} = 0{,}02557.$$

Упомянутая сейчасъ связь чиселъ проявляется при соединеніи ихъ въ суммы по два, по четыре и, въ особенности, по пяти. Вычисляя для этихъ 100, 50 и 40 комбинацій суммы квадратовъ ихъ отклоненій отъ

$$86{,}4, \quad 172.8 \quad \text{и} \quad 216,$$

мы получаемъ вмѣсто числа

$$5788.8$$

такія

$$3551{,}6, \quad 3089{,}2, \quad 1004,$$

послѣднее изъ которыхъ почти въ шесть разъ меньше числа 5788.8.