ON THE PROBLEM OF TWO SAMPLES*

By E. S. PEARSON and J. NEYMAN

Galton Laboratory, University College London and
Biometric Laboratory, Nencki Institute, Warsaw

I. The need for distinguishing between Criteria

Professor V. Romanovsky (1928) has contributed to a recent number of *Metron* an important paper entitled "On Criteria that two given Samples belong to the same Normal Population". The paper is divided into two parts; the first concerns the cases where the samples are compared with respect to one variable character only, while in the second it is supposed that measures on a number of characters have been made. We propose to consider here certain aspects of the first problem which is the more fundamental, although the development of analogous tests in the second problem is of considerable interest and appears to be novel.

The process of testing the hypothesis, H, that two observed samples, Σ_1 and Σ_2, have come from the same unknown normal population π consists (for Prof. Romanovsky) (1) in supposing H to be true, and (2) deducing the frequency distribution (determined by this hypothesis) in repeated random sampling, of certain functions of the sample means and standard deviations, which Professor Romanovsky calls the "belonging coefficients" or fully "coefficients of samples belonging to the same general population". The problem of deducing the frequency distribution of such coefficients is often a very difficult one and Professor Romanovsky has succeeded in reaching several new results, either expressing them in the form of series or by giving the moment coefficients of their distributions. This is an important achievement, but there lies at the very basis of the question another problem, which he does not appear to have fully considered, that of distinguishing between these coefficients or criteria and determining which is the most appropriate one to use in any given case. We have discussed elsewhere (Neyman & Pearson, 1928*a*, *b*; Neyman, 1929*a*) certain principles regarding the testing of hypotheses, which we believe to be intuitively sound. They are not, strictly speaking, mathematical results and may be rejected by those who do not believe in them. But we think that any reader who will follow us in examining Romanovsky's results will be convinced that some logical principle must be adopted in choosing between the various criteria.

Suppose that the two samples Σ_1 and Σ_2 have respectively

n_1 individuals with mean character $\bar{x}_1$ and standard deviation s_1,

n_2 individuals with mean character $\bar{x}_2$ and standard deviation s_2.

* Presented at the meeting on 10 February 1930 by M. Cz. Białobrzeski.

Romanovsky considers four alternative criteria for testing the hypothesis H, namely that Σ_1 and Σ_2 are random samples from the same normal population π. These are

$$\alpha=\frac{(\bar{x}_1-\bar{x}_2)^2}{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}; \quad (1) \qquad u=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)}}; \quad (2)$$

$$t=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{(n_1 s_1^2+n_2 s_2^2)}}\sqrt{\left\{\frac{n_1 n_2(n_1+n_2-2)}{n_1+n_2}\right\}}; \quad (3) \qquad \theta=\frac{s_2^2}{s_1^2}. \quad (4)$$

With regard to these criteria he gives the following results:

(*a*) The frequency distribution of α, in the form of a series, which however appears to be too complicated as it stands for practical purposes.

(*b*) The Normal Law as the limiting form of the frequency distribution of u. The proof is carried out by finding general expressions for the moment coefficients of u, and making use of the Tchebycheff–Markoff theorem regarding their limiting values (Markoff, 1912, p. 266). In the case of smaller samples these moment expressions might be used to fit some system of non-normal curves to the distribution.

(*c*) A new method is given of reaching the distribution of t, that originally given by R. A. Fisher (1925).

(*d*) The distribution of $\theta=s_2^2/s_1^2$ is given. This leads to the same test as used by Fisher for the comparison of two estimates of variance.*

As was pointed out in our earlier paper, in testing hypotheses two considerations must be kept in view, (1) we must be able to reduce the chance of rejecting a true hypothesis to as low a value as desired; (2) the test must be so devised that it will reject the hypothesis tested when it is likely to be false. In discussing the four criteria Romanovsky is thinking primarily of this first point, and in fact all his tests provide complete control of this source of error. Suppose that ϵ be any small positive number, that ξ be any statistical character of *both* samples Σ_1 and Σ_2, and that $\phi(\xi)$ represents the frequency distribution of ξ deduced on the hypothesis that both samples have been drawn at random from the same normal population. Choose now two numbers ξ_1 and ξ_2, such that

$$\int_{\xi_1}^{\xi_2}\phi(\xi)\,d\xi\leqslant\epsilon \qquad (5)$$

and accept the principle of rejecting the hypothesis H when, and when only, the value of ξ, say ξ_0, obtained from the observed samples is such that

$$\xi_1\leqslant\xi_0\leqslant\xi_2. \qquad (6)$$

As the chance of rejecting a true hypothesis H is the product of (1) the chance of dealing with a true hypothesis, and (2) the chance of obtaining a ξ_0 satisfying

* Fisher (1930, §41) uses the transformation $z=\frac{1}{2}\log\left\{\frac{n_2(n_1-1)\,s_2^2}{n_1(n_2-1)\,s_1^2}\right\}$ and has given certain tables of the z distribution.

(6), calculated on the assumption that H be true, the chance of rejecting a true hypothesis (when we reject only in cases where (6) holds good) will be certainly less than ϵ, that is as low as desired.

From this point of view, provided that (5) is satisfied, *any* values of ξ_1 and ξ_2 might be chosen, and further *any* statistical character ξ would be equally suitable. But is is clear when we examine the statistical tests in common use that a consciousness, if somewhat vaguely defined, of what we have termed the second source of error, has determined the particular characters ξ chosen as criteria and also the position of the limiting values ξ_1 and ξ_2 in the $\phi(\xi)$ distribution. For example, the first three characters α, u and t considered by Romanovsky depend upon the difference $\bar{x}_1-\bar{x}_2$. The reasoning which underlies this choice is presumably as follows. If the hypothesis H were false, then *probably* the means of the two sampled populations would differ, and this is likely to cause a large value of the difference $\bar{x}_1-\bar{x}_2$; therefore if this difference be large it is dangerous to accept the hypothesis. For this reason the numbers ξ_1 and ξ_2 are chosen in a manner slightly different to that indicated above, namely the hypothesis H is rejected when

$$\xi_0\leqslant\xi_1 \quad\text{or}\quad \xi_2\leqslant\xi_0 \qquad (7)$$

and the sum of the two integrals

$$\int_{-\infty}^{\xi_1}+\int_{\xi_2}^{+\infty}\phi(\xi)\,d\xi\leqslant\epsilon. \qquad (8)$$

But clearly the tests with α, u and t are not really sensitive to differences in population standard deviations. We may for instance have two pairs of samples in one of which s_1 and s_2 are almost equal, while in the second pair one is many times as great as the other; and yet both may provide the same value of t. That is to say, in certain cases the test would not distinguish what is otherwise obvious, that the standard deviations of the sampled populations may in one case be the same, but in the other can hardly be so. On the other hand the fourth criterion, θ, will distinguish differences between the population standard deviations, but is quite insensitive to differences between their means.

It would not be difficult to give examples of pairs of samples which would lead to contradictory conclusions about the hypothesis H when different tests are applied, and in such cases which conclusion is to be drawn? That unfavourable to the hypothesis? Not necessarily so, for given any two samples it is always possible to choose a statistical character ξ, or simply the numbers ξ_1 and ξ_2 so that the verdict based on such a ξ test would be unfavourable to H.*

* Neyman (1929*b*) has illustrated this point in testing the composite hypothesis that a single sample of n with mean $\bar{x}$ and standard deviation s has come from some normally distributed population with mean a. Here we should ordinarily use "Student's" t (or z) test, namely calculate $z=t\sqrt{(n-1)}=|x-a|/s$, refer to the appropriate tables and so find P_z, or the chance of obtaining in random sampling a value of z so large or larger than that observed. In n-dimensioned space z is the cotangent of the angle ψ between the vector representing the sample and the diagonal

$$x_1=x_2=\ldots=x_n;$$

It is for this reason that we have felt the need of finding a logical basis for the choice of one test rather than another. Owing to the fact that the control of the error involved in rejecting a true hypothesis can be obtained when *any* statistical character of the two samples is used, we may choose this character in such a way that it will minimise the danger of accepting the hypothesis when it is false. Of course this word "danger" must be clearly defined. We have assumed that the danger of the hypothesis being false can be measured by the "likelihood" of that hypothesis.* The justification of this principle rests upon the consideration of simple problems of testing hypotheses where an intuitive verdict may easily be reached. It appears that we are always inclined to reject a hypothesis H in cases where there are possible alternative hypotheses which make the probability of occurrence of the observed events (such as of drawing the samples Σ_1 and Σ_2) much greater than that determined by H. It was this idea that led to that definition of the likelihood of a hypothesis.

II. Application of the Method of Likelihood to the Problem

Let us try to consider the problem of Professor Romanovsky from this point of view, first making the conditions a little more precise. To commence with it is necessary to fix the set, Ω, of admissible hypotheses concerning the populations from which Σ_1 and Σ_2 have been drawn†. The most general assumption would be that these populations may be of any form whatsoever, normal or not normal. The less general assumption which we shall adopt is that Σ_1 and Σ_2 have come from *some* normal populations π_1 and π_2 having any values a_1 and a_2, σ_1 and σ_2 for their means and standard deviations. Then we wish to test the hypothesis H (which we have termed a composite hypothesis) that the populations π_1 and π_2 belong to the subset ω of Ω for which

$$a_1 = a_2, \quad \sigma_1 = \sigma_2, \tag{9}$$

that is to say, that π_1 and π_2 are identical.

The chance of drawing from populations π_1 and π_2 two samples Σ_1 and Σ_2 with values of the variables

$$x_1, x_2, \ldots, x_{n_1}; \quad x_{n_1+1}, x_{n_1+2}, \ldots, x_{n_1+n_2}$$

P_s is the integral of the normal spherical density field lying outside the hypercone with semi-vertical angle ψ.

If now we were to take as criterion not z, but ξ, the cotangent of the angle between the sample vector and the axis

$$\frac{x_1}{\alpha_1} = \frac{x_2}{\alpha_2} = \ldots = \frac{x_n}{\alpha_n}$$

the distribution of ξ in sampling would still follow Student's law. But for a given sample the value of P_ξ could be made to assume any value whatsoever lying between 0 and 1, by an appropriate choice of the direction cosines $\alpha_1, \alpha_2, \ldots, \alpha_n$, of the axis of the hypercone. A similar "nonsense" criterion could be found for the case of two samples.

* The term "likelihood" was introduced by R. A. Fisher (1922); we have somewhat extended the definition (see Neyman & Pearson, 1928*b*, pp. 264–5).

† The terminology is that explained in the second reference of the preceding footnote.

lying within the limits

$$x_1 - \tfrac{1}{2}h, x_1 + \tfrac{1}{2}h; \quad x_2 - \tfrac{1}{2}h, x_2 + \tfrac{1}{2}h; \text{ etc.}$$

will, as h tends to zero, be asymptotic to

$$C = \left(\frac{1}{\sigma_1\sqrt{(2\pi)}}\right)^{n_1} \left(\frac{1}{\sigma_2\sqrt{(2\pi)}}\right)^{n_2} \exp\left[-\left\{n_1 \frac{(\bar{x}_1 - a_1)^2 + s_1^2}{2\sigma_1^2} + n_2 \frac{(\bar{x}_2 - a_2)^2 + s_2^2}{2\sigma_2^2}\right\}\right] h^{n_1+n_2}. \tag{10}$$

It is easy to show that the maximum values of C associated with a pair of populations belonging to the set Ω is obtained when

$$a_1 = \bar{x}_1, a_2 = \bar{x}_2, \sigma_1 = s_1, \sigma_2 = s_2,$$

and hence
$$C(\Omega \max) = \left(\frac{1}{s_1\sqrt{(2\pi)}}\right)^{n_1} \left(\frac{1}{s_2\sqrt{(2\pi)}}\right)^{n_2} e^{-\frac{1}{2}(n_1+n_2)} h^{n_1+n_2}. \tag{11}$$

We must now find the maximum value of C associated with a population belonging to ω. This we may do by first noting that if $a_1 = a_2 = a$ (say) and $\sigma_1 = \sigma_2 = \sigma$ (say), we can write (10) as

$$C = \left(\frac{1}{\sigma\sqrt{(2\pi)}}\right)^{n_1+n_2} \exp\left[-\left\{(n_1+n_2)\frac{(\bar{x}_0 - a)^2 + s_0^2}{2\sigma^2}\right\}\right] h^{n_1+n_2}, \tag{12}$$

where $\bar{x}_0$ and s_0 are the mean and standard deviation obtained by combining the $n_1 + n_2$ variables of the samples Σ_1 and Σ_2. It now follows that the maximum chance occurs when

$$a = \bar{x}_0, \quad \sigma = s_0$$

and
$$C(\omega \max) = \left(\frac{1}{s_0\sqrt{(2\pi)}}\right)^{n_1+n_2} e^{-\frac{1}{2}(n_1+n_2)} h^{n_1+n_2}. \tag{13}$$

The expression for the likelihood of the composite hypothesis H that Σ_1 and Σ_2 have been randomly drawn from the same normal population is therefore

$$\lambda_H = \frac{C(\omega \max)}{C(\Omega \max)} = \left(\frac{s_1}{s_0}\right)^{n_1} \left(\frac{s_2}{s_0}\right)^{n_2}. \tag{14}$$

If this ratio be small, it means that whatsoever be the hypothetical population π from which we may suppose Σ_1 and Σ_2 have been drawn, the probability of obtaining these samples from π is much smaller than the probability of getting them from some two *different* normal populations. On the other hand, if the λ_H of (14) is a number approaching unity, then there are certainly some normal populations π such that the probability of drawing Σ_1 and Σ_2 from π is not so very much less than that of drawing them from two different populations. Speaking roughly, in the first event there are no normal populations from which both samples could easily have come, while in the second such populations do exist. λ_H is therefore the criterion suggested for testing the hypothesis H.

It will be noted that

$$s_0^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2} + \frac{n_1 n_2}{(n_1+n_2)^2}(\bar{x}_1 - \bar{x}_2)^2 \tag{15}$$

and therefore if we write

$$t=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{(n_1s_1^2+n_2s_2^2)}}\sqrt{\left\{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}\right\}} \tag{16}$$

and

$$\theta=s_2^2/s_1^2, \tag{17}$$

we obtain

$$\lambda=(n_1+n_2)^{\frac{1}{2}(n_1+n_2)}\theta^{\frac{1}{2}n_2}(n_1+n_2\theta)^{-\frac{1}{2}(n_1+n_2)}\left(1+\frac{t^2}{n_1+n_2-2}\right)^{-\frac{1}{2}(n_1+n_2)}, \tag{18}$$

that is to say λ_H can be expressed as a function of two of the four coefficients given by Romanovsky. This is a point of considerable interest, because the tests associated with t and θ (which are the "Student" and R. A. Fisher tests) are in fact those suggested by measuring the likelihood not of the full hypothesis H, but of certain modified hypotheses H_1 and H_2. We shall first proceed to consider this point.

In all three cases we assume that Σ_1 and Σ_2 have come from some two normal populations, but

(1) The hypothesis H_1 is that the samples have come from unknown normal populations *with the same variance*, but with means having any values a_1 and a_2 whatsoever. Here θ is the appropriate criterion.

(2) While if it be assumed that the populations besides being normal have the same variance, then the hypothesis H_2 is that the means in these populations are the same. Here t is the appropriate criterion.

The deduction of the t-test from the principle of likelihood has already been discussed by us (Neyman & Pearson, 1928*a*, pp. 206–7). It was shown that, if t is as defined in (16), then the likelihood of the hypothesis H_2 is

$$\lambda_{H_2}=\left(1+\frac{t^2}{n_1+n_2-2}\right)^{-\frac{1}{2}(n_1+n_2)} \tag{19}$$

and as "Student" has shown, the frequency distribution of t in repeated samples Σ_1 and Σ_2 from a common normal population is

$$f(t)=\frac{\Gamma(\frac{1}{2}[n_1+n_2-1])}{\sqrt{\{\pi(n_1+n_2-2)\}}\,\Gamma(\frac{1}{2}[n_1+n_2-2])}\left(1+\frac{t^2}{n_1+n_2-2}\right)^{-\frac{1}{2}(n_1+n_2-1)} \tag{20}$$

λ_{H_2} has its maximum value of unity when $t=0$ or $\bar{x}_1=\bar{x}_2$, and decreases as $t\to\pm\infty$. The chance of obtaining in sampling a value of λ_{H_2} less than that observed may therefore be obtained from the probability integral of (20).

Let us examine the likelihood of hypothesis H_1. Now Ω is the set of all possible pairs of normal populations with means a_1 and a_2, and standard deviations σ_1 and σ_2, and $C(\Omega\max)$ has the value given in (11). But ω is the subset of populations in which $\sigma_1=\sigma_2$ while a_1 and a_2 have any values. We must therefore choose a_1, a_2 and a common value of $\sigma_1=\sigma_2=\sigma$ (say) as to maximise

$$C=\left(\frac{1}{\sigma\sqrt{(2\pi)}}\right)^{n_1+n_2}\exp\left[-\frac{\{n_1[(\bar{x}_1-a_1)^2+s_1^2]+n_2[(\bar{x}_2-a_2)^2+s_2^2]\}}{2\sigma^2}\right]h^{n_1+n_2}. \tag{21}$$

It is easily found that the values required are

$$a_1=\bar{x}_1,\quad a_2=\bar{x}_2,\quad \sigma^2=\frac{n_1s_1^2+n_2s_2^2}{n_1+n_2},$$

$$C(\omega\max)=\left(\frac{1}{\sqrt{(2\pi)}}\right)^{n_1+n_2}\left(\frac{n_1s_1^2+n_2s_2^2}{n_1+n_2}\right)^{-\frac{1}{2}(n_1+n_2)}e^{-\frac{1}{2}(n_1+n_2)}h^{n_1+n_2}, \tag{22}$$

whence

$$\lambda_{H_1}=\frac{C(\omega\max)}{C(\Omega\max)}=(n_1+n_2)^{\frac{1}{2}(n_1+n_2)}\frac{s_1^{n_1}s_2^{n_2}}{(n_1s_1^2+n_2s_2^2)^{\frac{1}{2}(n_1+n_2)}}$$

$$=(n_1+n_2)^{\frac{1}{2}(n_1+n_2)}\theta^{\frac{1}{2}n_2}(n_1+n_2\theta)^{-\frac{1}{2}(n_1+n_2)}. \tag{23}$$

Further, as Romanovsky has shown, the distribution of θ in repeated samples Σ_1 and Σ_2 from normal populations with a common variance is

$$\phi(\theta)=\frac{n_1^{\frac{1}{2}(n_1-1)}n_2^{\frac{1}{2}(n_2-1)}}{\Gamma(\frac{1}{2}[n_1-1])\,\Gamma(\frac{1}{2}[n_2-1])}\Gamma\left(\frac{n_1+n_2-2}{2}\right)\theta^{\frac{1}{2}(n_2-3)}(n_1+n_2\theta)^{-\frac{1}{2}(n_1+n_2-2)}. \tag{24}$$

λ_{H_1} has its maximum value of unity when $\theta=1$, or $s_1=s_2$ and decreases as θ tends to zero or infinity, that is to say, as the difference between s_1 and s_2 increases.

It will be noticed at once from (18), (19) and (23) that

$$\lambda_H=\lambda_{H_1}\lambda_{H_2}. \tag{25}$$

We are now able to see more clearly the interpretation of λ_H. The likelihood can only attain its maximum value of unity when both $\theta=1$ and $t=0$, or $s_1=s_2$ and $\bar{x}_1=\bar{x}_2$. It will decrease towards zero when

(*a*) $\theta\to 0$ or s_2 becomes small compared with s_1

or (*b*) $\theta\to\infty$ or s_1 becomes small compared with s_2

or (*c*) $|t|\to\infty$ or $|\bar{x}_1-\bar{x}_2|$ increases compared with

$$\sqrt{\left\{\frac{n_1s_1^2+n_2s_2^2}{n_1+n_2-2}\cdot\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right\}}.$$

It is clear that, on the one hand, even if $\bar{x}_1=\bar{x}_2$ or $\lambda_{H_2}=1$ we should be unable to accept H if s_1 differed considerably from s_2 (or λ_{H_1} was small). While on the other hand if $s_1=s_2$ (or $\lambda_{H_1}=1$) we should feel that the populations were not the same if $[\bar{x}_1-\bar{x}_2|$ were large compared to

$$\sqrt{\left\{\frac{n_1s_1^2+n_2s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right\}}$$

which is an estimate based on the sample variances of the standard error of the difference in means. Thus the criterion λ_H is more crucial than either λ_{H_1} or λ_{H_2} taken separately, although of course in many problems the hypothesis to be tested will present itself in one of these common forms H_1 or H_2.*

* For example, (*a*) two different angles a_1 and a_2 are measured n_1 and n_2 times respectively by the same person using the same instrument. It may then be supposed that the precision of measurements will be the same in both cases, $\sigma_1=\sigma_2$, but we may wish to test hypothesis H_2, namely that

From another point of view it is interesting to note the form of λ_H given in (14); we see that the hypothesis H becomes more and more improbable as either of the sample standard deviations decreases compared to the standard deviation of the pooled samples. This again is in agreement with our intuition.

These simple considerations indeed suggest that λ_H is a reasonable criterion to use in measuring the danger of accepting a false hypothesis, and it appears to satisfy our intuitive requirements in a way which neither α or u, θ, or t are able to do*. But if we accept the criterion suggested by the method of likelihood it is still necessary to determine its sampling distribution in order to control the error involved in rejecting a true hypothesis, because a knowledge of λ alone is not adequate to insure control of this error. We cannot for example say in general that if $\lambda \leqslant \lambda_0 = 0{\cdot}01$, we should be justified in rejecting the hypothesis. In order to fix a limit between "small" and "large" values of λ we must know how often such values appear when we deal with a true hypothesis. That is to say we must have knowedge of P_{λ_0}, the chance of obtaining $\lambda \leqslant \lambda_0$ in the case where the hypothesis tested is true. The frequency distribution of λ differs according to the size of samples and the nature of the hypothesis tested, and it may well happen that the modal value of λ is in the neighbourhood of zero. For the cases of H_1 and H_2 the distributions (24) and (20) of θ and t provide what is required, but for H the position is not so simple.

We know that the frequency function for simultaneous variations of $\bar{x}_1, \bar{x}_2, s_1$ and s_2 is

$$C_0 s_1^{n_1-2} s_2^{n_2-2} \exp -\left\{\frac{n_1(\bar{x}_1-a)^2+n_2(\bar{x}_2-a)^2+n_1 s_1^2+n_2 s_2^2}{2\sigma^2}\right\}. \tag{26}$$

The value of C_0 is

$$C_0=\frac{4 n_1^{\frac{1}{2}n_1} n_2^{\frac{1}{2}n_2}}{\{\pi\sqrt{(2\sigma)}\}^{\frac{1}{2}(n_1+n_2)}\Gamma(\frac{1}{2}[n_1-1])\,\Gamma(\frac{1}{2}[n_1-1])}.$$

Now transform the variables to

$$\left.\begin{aligned} \bar{x}_2' &= \bar{x}_2 - a; \quad \xi=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{(n_1 s_1^2+n_1 s_2^2)}}\sqrt{\left(\frac{n_1 n_2}{n_1+n_2}\right)}\\ \theta &= \frac{s_2^2}{s_1^2}; \quad q=n_1 s_1^2+n_2 s_2^2. \end{aligned}\right\} \tag{27}$$

Then it follows that

$$s_1^2=\frac{q}{n_1+n_2\theta}, \quad s_2^2=\frac{\theta q}{n_1+n_2\theta}, \tag{28}$$

while it is not difficult to show that

$$n_1(\bar{x}_1-a)^2+n_2(\bar{x}_2-a)^2=\{\sqrt{(n_1+n_2)}\,\bar{x}_2'+\xi\sqrt{q}\,\sqrt{(n_1/n_2)}\}^2+\xi^2 q. \tag{29}$$

$a_1=a_2$. (b) Some two angles a_1 and a_2 are measured several times by two observers or by the same observer with two different instruments. The question arises as to whether the precision of the two sets of measurements is the same. Does $\sigma_1=\sigma_2$? This is the test of hypothesis H_1. No assumption is made about the values of a_1 and a_2.

* Besides the paper in *Metron* already referred to, Professor Romanovsky published in Russian a short Note on the same subject in which there is a sentence stating that the tests like t and θ concern the hypothesis H from two different points of view. Evidently the author felt himself that really he was testing not the hypothesis H, but some other hypotheses.

Further the Jacobian of the transformation gives

$$\left|\frac{\partial(\bar{x}_1,\bar{x}_2,s_1,s_2)}{\partial(\bar{x}_2',\xi,q,\theta)}\right|=\frac{1}{4}\sqrt{\left(\frac{n_1+n_2}{n_1 n_2}\right)}\,\theta^{-\frac{1}{2}}(n_1+n_2\theta)^{-1}q^{\frac{1}{2}}. \tag{30}$$

Hence the probability law of $\bar{x}_2', \xi, q, \theta$ is

$$C_0\frac{1}{4}\sqrt{\left(\frac{n_1+n_2}{n_1 n_2}\right)}\,q^{\frac{1}{2}(n_1+n_2-3)}\,\theta^{\frac{1}{2}(n_2-3)}(n_1+n_2\theta)^{-\frac{1}{2}(n_1+n_2-2)}$$
$$\exp -\frac{\{\bar{x}_2'\sqrt{(n_1+n_2)}+\xi\sqrt{q}\,\sqrt{(n_1/n_2)}\}^2}{2\sigma^2}\,e^{-q(1+\xi^2)/2\sigma^2}. \tag{31}$$

We may now integrate for $\bar{x}_2'$ between the limits $-\infty$ and $+\infty$ and shall merely have in place of the $\bar{x}_2'$ term a factor $\{2\pi/(n_1+n_2)\}^{\frac{1}{2}}\sigma$. Hence the probability law of ξ, q and θ is

$$C_0\frac{\sigma}{4}\sqrt{\frac{2\pi}{n_1 n_2}}\,\theta^{\frac{1}{2}(n_2-3)}(n_1+n_2\theta)^{-\frac{1}{2}(n_1+n_2-2)}q^{\frac{1}{2}(n_1+n_2-3)}\exp -\left\{\frac{q(1+\xi^2)}{2\sigma^2}\right\}. \tag{32}$$

To integrate for q between the limits 0 and $+\infty$ we write $u=q(1+\xi^2)(2\sigma^2)^{-1}$ and find the probability law of ξ and θ

$$F_1(\xi,\theta)=C_0\sigma^{n_1+n_2}2^{\frac{1}{2}(n_1+n_2-4)}\sqrt{\left(\frac{\pi}{n_1 n_2}\right)}\,\theta^{\frac{1}{2}(n_2-3)}(n_1+n_2\theta)^{-\frac{1}{2}(n_1+n_2-2)}$$
$$\times(1+\xi^2)^{-\frac{1}{2}(n_1+n_2-1)}\int_0^\infty u^{\frac{1}{2}(n_1+n_2-3)}e^{-u}\,du,$$

which on substituting $\Gamma(\frac{1}{2}[n_1+n_2-1])$ for the u-integral and using the value for C_0 given above results in

$$\frac{n_1^{\frac{1}{2}(n_1-1)}n_2^{\frac{1}{2}(n_2-1)}\Gamma(\frac{1}{2}[n_1+n_2-1])}{\sqrt{\pi}\,\Gamma(\frac{1}{2}[n_1-1])\,\Gamma(\frac{1}{2}[n_2-1])}\,\theta^{\frac{1}{2}(n_2-3)}(n_1+n_2\theta)^{-\frac{1}{2}(n_1+n_2-2)}(1+\xi^2)^{-\frac{1}{2}(n_1+n_2-1)}. \tag{33}$$

If now we substitute for ξ, $t=\sqrt{(n_1+n_2-2)}\,\xi$, we see on comparison with (20) and (24) that

$$F(t,\theta)=F_1\left(\frac{t}{\sqrt{(n_1+n_2-2)}},\theta\right)=f(t)\,\phi(\theta). \tag{34}$$

That is to say the simultaneous distribution of t and θ is identical with the product of their independent distributions, or in pairs of random samples Σ_1 and Σ_2 from a common normal population, the two coefficients t and θ are completely uncorrelated. This result (34) combined with the previous result (25) is of considerable interest. $F(t,\theta)$ may be represented by a two-dimension density field; for a given value of θ the change in t follows a Pearson Type VII law; for a given t the change in θ follows a Type VI law; for a given t the change in θ follows a Type VI law. As there is no correlation, the array distributions are homoscedastic and the same as the marginal distributions.

As shown by (18) the contours of constant λ_H are closed oval curves in the field, not corresponding exactly to the contours of equal density, but tending to these as n_1 and n_2 are increased. The chance of obtaining in pairs of samples a value of λ_H less than any given amount, P_{λ_H}, is the integral of the density represented by (33) taken outside the corresponding λ_H-contour represented by (18). We shall discuss below an approximate method of obtaining this probability integral.

A limiting case of the problem we have considered will occur if we make $n_2 \to +\infty$, but keep n_1 finite. The second sample then becomes indistinguishable from an infinite population, so that $\bar{x}_2 = a$ and $s_2 = \sigma$. In fact we are now testing a simple hypothesis, namely that a sample of $n(=n_1)$ with mean $\bar{x}(=\bar{x}_1)$ and standard deviation $s(=s_1)$ has been drawn from a normal population with mean and standard deviation of a and σ respectively. It is not difficult to show that the limiting form as $n_2 \to \infty$ of λ_H becomes

$$\lambda'_H = \left(\frac{s}{\sigma}\right)^n \exp\left[-\tfrac{1}{2}n\left\{\frac{(\bar{x}-a)^2+s^2}{\sigma^2}-1\right\}\right]. \tag{35}$$

This is the problem we have considered elsewhere (1928*a*, pp. 187–9)*. In that case $m = \bar{x} - a$ and s^2 are two completely independent criteria and their sampling distribution is represented by a density field in which the distribution of m for a given s is a normal curve and that of s for a given m a Pearson Type III curve. The contours of λ'_H were also closed oval curves tending to correspond to the contours of equal density as n was increased. In our paper we gave diagrams of these contours and also tables of P_λ, the chance of drawing a sample with λ less than any given value. These tables form what might be termed the marginal distributions, corresponding to $n_2 \to \infty$ or $n_1 \to \infty$, of the tables required in our present problem. They involved considerable computation using quadratures; we shall be content in the present paper with a more approximate solution of the problem of finding P_{λ_H} for the case of two samples.

III. The Moments of the Distribution of λ_H and Approximations to P_{λ_H}.

It is not difficult to find the moment coefficients of the distribution of λ_H obtained when repeated pairs of random samples are drawn from the same normal population. We shall first obtain these coefficients and then outline the stages which have led to the provisional tables connecting λ_H and P_{λ_H} which we give at the end of this paper. A more detailed analysis of the problem is required, but this we must leave for a later paper.

We have the two relations

$$\lambda_H = (n_1+n_2)^{\frac{1}{2}(n_1+n_2)}\theta^{\frac{1}{2}n_2}(n_1+n_2\theta)^{-\frac{1}{2}(n_1+n_2)}(1+\xi^2)^{-\frac{1}{2}(n_1+n_2)} \tag{36}$$

and

$$F_1(\xi,\theta) = C_0\theta^{\frac{1}{2}(n_2-3)}(n_1+n_2\theta)^{-\frac{1}{2}(n_1+n_2-2)}(1+\xi^2)^{-\frac{1}{2}(n_1+n_2-1)}, \tag{37}$$

* The hypothesis was there termed Hypothesis A. Equation (35) above corresponds to (xix) of our previous paper.

where C_0 is the constant term given in (33) above. Denote by μ'_k the kth moment-coefficient of λ_H about zero, then

$$\mu'_k = \int_0^\infty \int_{-\infty}^{+\infty} F_1(\xi,\theta)\lambda_H^k\, d\theta\, d\xi$$
$$= C_0(n_1+n_2)^{\frac{1}{2}(n_1+n_2)k}\int_0^{+\infty}\theta^{\frac{1}{2}[(k+1)n_2-3]}(n_1+n_2\theta)^{-\frac{1}{2}[(k+1)(n_1+n_2)-2]}\,d\theta$$
$$\times \int_{-\infty}^{+\infty}(1+\xi^2)^{-\frac{1}{2}[(k+1)(n_1+n_2)-1]}\,d\xi = C_0(n_1+n_2)^{\frac{1}{2}(n_1+n_2)k} I_\theta I_\xi.$$

The integrals I_θ and I_ξ may be calculated separately. Writing $u = n_1/(n_1+n_2\theta)$ we obtain a Beta Function for the first, or

$$I_\theta = n_1^{-\frac{1}{2}[(k+1)n_2-1]}n_2^{-\frac{1}{2}[(k+1)n_2-1]}B(\tfrac{1}{2}[(k+1)n_1-1], \tfrac{1}{2}[(k+1)n_2-1]) \tag{38}$$

and writing $v = (1+\xi^2)^{-1}$

$$I_\xi = B(\tfrac{1}{2}[(k+1)(n_1+n_2)-2], \tfrac{1}{2}). \tag{39}$$

From these results it follows that

$$\mu'_k = \left(\frac{(n_1+n_2)^{n_1+n_2}}{n_1^{n_1}n_2^{n_2}}\right)^{\frac{1}{2}k}\frac{\Gamma(\frac{1}{2}[(k+1)n_1-1])\,\Gamma(\frac{1}{2}[(k+1)n_2-1])}{\Gamma(\frac{1}{2}[(k+1)(n_1+n_2-1)])}$$
$$\times\frac{\Gamma(\frac{1}{2}[n_1+n_2-1])}{\Gamma(\frac{1}{2}[n_1-1])\,\Gamma(\frac{1}{2}[n_2-1])}. \tag{40}$$

The limiting value of (40) in two cases is of interest.

(1) *When n_1 and n_2 both become very large.* Using the first approximation given by Stirling's Formula or,

$$\Gamma(x) \sim x^{x-\frac{1}{2}}e^{-x}\sqrt{(2\pi)} \tag{41}$$

it is found that as n_1 and $n_2 \to \infty$

$$\mu'_k \to 1/(k+1) \tag{42}$$

uniformly whatever be k. We conclude that the frequency distribution of λ tends at the same time to a limiting form having its kth moment equal to $(k+1)^{-1}$. It is the so called Rectangular Distribution, or $f(\lambda) = \text{constant}$. We therefore appear justified in assuming that when comparing large samples, the frequency distribution of λ_H will tend to this form, or P_{λ_H} will approach λ_H in value.

(2) *When n_2 becomes very large, n_1 remaining finite.* Making $n_2 \to \infty$ in (40) and using (41) we find that

$$\mu'_k \to \frac{\Gamma(\frac{1}{2}[(k+1)n_1-1])}{\Gamma(\frac{1}{2}[n_1-1])}(k+1)^{-\frac{1}{2}[(k+1)n_1]}\left(\frac{2e}{n_1}\right)^{\frac{1}{2}kn_1}, \tag{43}$$

an expression giving the moment coefficients of the distribution of λ whose probability integral we have tabled in our earlier paper.

In order to obtain some appreciation of the rapidity with which the distribution of λ_H in the two sample problem tends to the rectangular form, we have calculated from (40) the frequency constants: mean λ, standard deviation of

λ, β_1 and β_2, for a variety of values of n_1 and n_2*. These are given in Table I. For the rectangular distribution these constants have values

Mean $=0\cdot5$; Standard deviation $=1/\sqrt{12}=0\cdot288675$; $\beta_1=0$; $\beta_2=1\cdot8$.

We have also calculated certain values of (43). All constants approach the values for the rectangular distribution as n_1 and n_2 are increased, but without further analysis it is not possible to say at what point it would be justifiable to use the equality $P_{\lambda_H}=\lambda_H$. We propose here only to set out briefly the steps in reasoning which have led us to the rough Tables II and III given below.

(*a*) As λ_H must lie between 0 and 1 we assume that its distribution can be represented approximately by the law

$$f(\lambda)=\frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)}\lambda^{p-1}(1-\lambda)^{q-1}. \tag{44}$$

The rectangular distribution is a special case of this, arising when $p=q=1$.

(*b*) The p and q of (44) may be expressed in terms of the mean, m_1, and the variance, m_2, of the distribution; namely

$$p=\frac{m_1}{m_2}\{m_1(1-m_1)-m_2\},\quad p=\frac{1-m_1}{m_2}\{m_1(1-m_1)-m_2\}. \tag{45}$$

(*c*) Substituting in (45) the true means and standard deviations of the distributions given in the third and fourth columns of Table I, that is putting $m_1=\text{Mean }\lambda_H$, $m_2=\mu_2'-(\text{Mean }\lambda_H)^2$ we have calculated p and q in each case.

(*d*) The curves represented by (44) have the following values for β_1 and β_2,

$$\left.\begin{aligned}\beta_1&=\frac{4(p-q)^2(p+q+1)}{pq(p+q+2)^2},\\ \beta_2&=\frac{3\beta_1(p+q+2)+6(p+q+1)}{2(p+q+3)}.\end{aligned}\right\} \tag{46}$$

These have been calculated using the p's and q's of (*c*), and the resulting values are given in the 7th and 8th columns of Table I.

(*e*) It will be seen that these latter values agree to the 4th decimal place very closely with the true values of β_1 and β_2 for the λ_H distribution calculated from the moment coefficients of equations (40) and (43) and given in the 5th and 6th columns of the table. This suggests that the curves (44) may give a reasonable fit to the true λ_H distribution, although a more exact confirmation is clearly required in the critical region near $\lambda_H=0$. A partial check on this point is described below.

(*f*) Using these curves (44) we have computed for several different sizes of the samples the values of λ_H for which

$$P_{\lambda_H}=\int_0^{\lambda_H} f(\lambda)\,d\lambda=0\cdot05\quad\text{and}\quad 0\cdot01.$$

* Mean $\lambda=\mu_1'$, Standard deviation $=\sqrt{(\mu_2'-\mu_1'^2)}$, and if μ_2, μ_3 and μ_4 are the second, the third and fourth moment-coefficients of λ referred to the mean,

$$\beta_1=\mu_3^2/\mu_2^3,\quad \beta_2=\mu_4/\mu_2^2.$$

These are given in Tables II and III. The computation was rendered easy because in all cases the value of q differed only slightly from unity, and therefore in the neighbourhood of $\lambda_H=0$ we obtain approximately from (44)

$$P_{\lambda_H}=\frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}\frac{\lambda^p}{p}. \tag{47}$$

TABLE I

Moment Coefficients of λ_H Distribution

		True values from equations (40) and (43)				Values from equations (46)	
n_1	n_2	Mean λ	σ_λ	β_1	β_2	β_1	β_2
5	5	·4222	·2986	·0787	1·8257	·0779	1·8252
5	10	·4403	·2973	·0451	1·8001	·0450	1·8001
5	20	·4459	·2969	·0365	1·7942	·0367	1·7943
5	50	·4477	·2968	·0339	1·7924	·0342	1·7926
5	∞	·4481	·2968	·0334	1·7920	·0338	1·7922
10	10	·4634	·2946	·0166	1·7863	·0165	1·7863
10	20	·4718	·2935	·0098	1·7855	·0098	1·7855
10	50	·4749	·2931	·0077	1·7856	·0077	1·7856
10	∞	·4757	·2930	·0072	1·7857	·0072	1·7857
20	20	·4823	·2918	·00383	1·7879	·00382	1·7880
20	50	·4869	·2911	·00208	1·7901	·00208	1·7901
20	∞	·4882	·2909	·00168	1·7908	·00169	1·7908
50	50	·4930	·2900	·00059	1·7940	·00058	1·7941
50	∞	·4954	·2896	·00026	1·7959	·00026	1·7959
∞	∞	·5000	·2887	·00000	1·8000	·0000	1·8000

(*g*) For the limiting cases n_2 (or n_1) $=\infty$ we may compare the values of λ_H corresponding to $P_\lambda=0\cdot05$ and $0\cdot01$ computed in this approximate manner with those obtained from the tables in our earlier paper (Neyman & Pearson, 1928*a*, pp. 238–40). The latter were obtained by a quadrature of the density field lying inside the oval λ contours in the (m, s) field i.e. by a completely independent and exact method. They are given in brackets in the marginal columns of Tables II and III, and will be seen to agree very closely with the results of our present approximation.

(*h*) As a further check on the accuracy of this method of approximation, the case $n_1=n_2=5$ was taken and the integration of $F(t,\theta)$ outside certain of the λ_H contours was carried out by using quadratures. It was found that,

for $P_{\lambda_H}=0\cdot01$, $\lambda_H=0\cdot00193$ against the approximate value of $0\cdot00186$;

for $P_{\lambda_H}=0\cdot05$, $\lambda_H=0\cdot0169$ against the approximate value of $0\cdot0167$.

The former values are given in brackets in the cells corresponding to $n_1=n_2=5$ of Tables II and III.

(*k*) As far as can be judged from the frequency constants the distribution of λ for the limiting case $n_2=\infty$ and also for the special case $n_1=n_2=5$ are of the same general type as in the other cases. The agreement therefore, described in (*g*) and (*h*) above, between the true values of λ found by quadrature and those found by using the curves (44), suggests that Tables II and III may be taken throughout as giving adequate approximations to the values of λ_H corresponding to $P_\lambda=0{\cdot}05$ and 0·01.

TABLE II

Values of λ_H giving $P_{\lambda H}=0{\cdot}05$

n_2 \ n_1	5	10	20	50	∞
5	·0167 (·0169)	·0222	·0241	·0247	·0248 (·0247)
10	·0222	·0312	·0349	·0364	·0368 (·0367)
20	·0241	·0349	·0401	·0425	·0432 (·0431)
50	·0247	·0364	·0425	·0459	·0473 (·0474)
∞	·0248 (·0247)	·0368 (·0367)	·0432 (·0431)	·0473 (·0474)	·0500

TABLE III

Values of λ_H giving $P_{\lambda_H}=0{\cdot}01$

n_2 \ n_1	5	10	20	50	∞
5	·0019 (·0019)	·0029	·0033	·0034	·0034 (·0033)
10	·0029	·0048	·0058	·0061	·0062 (·0062)
20	·0033	·0058	·0071	·0078	·0080 (·0079)
50	·0034	·0061	·0078	·0088	·0092 (·0092)
∞	·0034 (·0033)	·0062 (·0062)	·0080 (·0079)	·0092 (·0092)	·0100

IV. AN ILLUSTRATIVE EXAMPLE

In our previous paper (loc. cit. p. 202) we took as an example the variations in Cephalic Index (breadth to length ratio × 100) measured on each of two series of 10 human skulls, namely

Series 1,

74·1; 77·7; 74·4; 74·0; 73·8; 72·2; 75·2; 78·2; 77·1; 78·4

$$n_1=10;\quad \bar{x}_1=75{\cdot}15;\quad s_1=2{\cdot}059.$$

Series 2,

66·7; 69·4; 67·8; 73·2; 79·3; 80·7; 64·9; 82·2; 72·4; 78·1

$$n_2=10;\quad \bar{x}_2=73{\cdot}47;\quad s_2=5{\cdot}942.$$

We then inquired whether it was likely, as far as the cephalic index was concerned that these two sets of skulls could be random samples from a population in which the mean cephalic index was 75·06 and the standard deviation 2·68, and concluded that it was very probable in the first case ($P_\lambda=0{\cdot}504$), but highly improbable in the second ($P_\lambda<0{\cdot}0001$). We were then testing two "simple" hypotheses.

We may now proceed to test the "composite" hypothesis that these two samples have come from the same population, without specifying what that population may be, except that it will be assumed that the distribution of cephalic index does not differ so much from normality as to invalidate the test.* On combining the two samples it is found that $s_0=4{\cdot}562$, and hence

$$\lambda_H=(s_1 s_2/s_0^2)^{10}=0{\cdot}00492.$$

Table III indicates that this value corresponds very closely to $P_{\lambda_H}=0{\cdot}01$; that is to say only once in a hundred times should we expect our criterion to have as low or a lower value were the hypothesis tested true. We should therefore conclude that it was very unlikely that the two series of skulls came from the same population. The values of t and θ for this pair of samples are

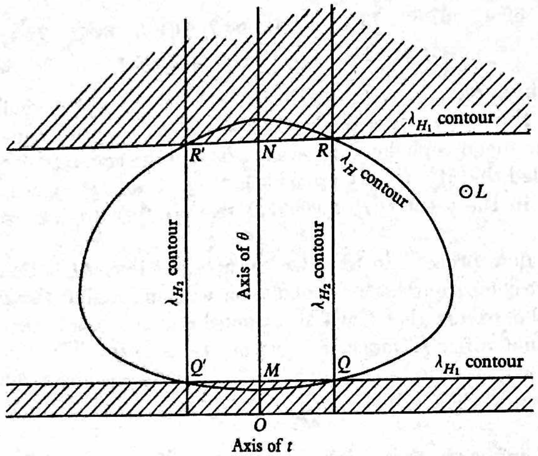
$$t=0{\cdot}973,\quad \theta=8{\cdot}328.$$

The point $R(t,\theta)$ representing the samples has been plotted in the accompanying diagram, and also a curve representing (though not drawn to scale) the member of the family of contours (18) for which $\lambda_H=0{\cdot}00492$. Then P_{λ_H} is the integral of the function $F(t,\theta)$ of (33) and (34) taken throughout the region of the field lying outside this contour.

Let us also consider the hypotheses H_1 and H_2. Equation (23) gives $\lambda_{H_1}=0{\cdot}00822$; this is constant for sample points lying not only on the line $R'NR$ through R, parallel to the axis of t, but also on a second parallel $Q'MQ$. The latter corresponds to certain pairs of samples in which s_1 is greater than s_2, and $\theta=0{\cdot}120$. $P_{\lambda_{H_1}}$ is the integral of $F(t,\theta)$ over (*a*) the region for which

* The sensitiveness of the test to deviations from normality will require further consideration.

$\theta \geqslant 8{\cdot}328$ and (*b*) the region for which $\theta \leqslant 0{\cdot}120$, that is to say over the two shaded areas of the diagram. It is found that $P_{\lambda_{H_1}} = 0{\cdot}0042$*. We should therefore conclude that it was extremely unlikely that the two samples came from populations with a common standard deviation, and as the test of hypothesis H_2 is based on the assumption that the populations sampled have a commmon standard deviation, we should hardly go on to examine this. If however we had some *a priori* grounds for believing that $\sigma_1 = \sigma_2$, and that therefore the observed difference between s_1 and s_2 was just a very abnormal chance fluctuation, then we might



continue and examine hypothesis H_2, and should find $\lambda_{H_2} = 0{\cdot}599$, and entering Student's (1925) Tables with $t = 0{\cdot}973$, $n = 18$, $P_{\lambda_{H_2}} = 0{\cdot}343$. That is to say we should conclude that there was nothing exceptional in the difference in mean cephalic indexes. It will be noted that

$$\lambda_H = 0{\cdot}00492 = \lambda_{H_1}\lambda_{H_2} = 0{\cdot}00822 \times 0{\cdot}599$$

as follows from (25).

It may appear at first to be illogical that while $P_{\lambda_H} = 0{\cdot}01$, $P_{\lambda_{H_1}} = 0{\cdot}0042$. But it must be remembered that in testing hypothesis H_1 we are only questioning whether it is likely that the population standard deviations are the same. Consequently a pair of samples corresponding to a point (t, θ) at L in the diagram would be more favourable to the hypothesis H_1 than the observed pair at R, and the region of the field in the neighbourhood of L is excluded in obtaining the

* Using Fisher's transformation and the tables referred to in the footnote on page 100 above we can only find that the integral over each shaded area is less than 0·01. The value 0·0042 was found by making the transformation $u = (1+\theta)^{-1}$ and using unpublished tables of the Incomplete Beta Function Integral.

$P_{\lambda_{H_1}}$ integral. But in testing the hypothesis H we are examining the two samples as a whole; the point L corresponds it is true to a lower value of θ but to a larger value of t than does R. It lies on an outer likelihood (λ_H) contour to R, and the region in which it lies is included in the P_{λ_H} integral. If the essential difference in the nature of the hypotheses H and H_1 is understood, it will be seen that there is no inconsistency in the fact that for the pair of samples used in our illustration $P_{\lambda_H} > P_{\lambda_{H_1}}$.

REFERENCES

FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans.* A, **222**, 309–68.

FISHER, R. A. (1925). Applications of 'Student's' distribution. *Metron*, **5** (3) 90–104.

FISHER, R. A. (1930). *Statistical Methods for Research Workers* (3rd edition). Edinburgh: Oliver and Boyd.

MARKOFF, A. A. (1912). *Wahrscheinlichkeitsrechnung*. Teubner (translation of the Russian edition of 1908).

NEYMAN, J. (1929*a*). Contribution to the theory of certain test criteria. *Bull. Int. Statist. Inst.* **24**, 44–88.

NEYMAN, J. (1929*b*). Contributions à la théorie de vraisemblance des hypothèses statistiques. *Rev. Trimestrielle Statist. de la République Polonaise*, **6** (4), 1–28.

NEYMAN, J. & PEARSON, E. S. (1928*a*, *b*). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, **20**, A. (*a*) Part I, 175–240; (*b*) Part II, 263–94.

ROMANOVSKY, V. (1928). On criteria that two given samples belong to the same normal population. *Metron*, **7** (3), 3–46.

STUDENT. (1925). New tables for testing the significance of observations. *Metron*, **5** (3), 105–11.