Telephone / Telegrams | Harpenden 621 (2 lines)

Railway Station: Harpenden (L.M.S.)

ROTHAMSTED EXPERIMENTAL STATION

(LAWES AGRICULTURAL TRUST)

Director: SIR E. JOHN RUSSELL, D.Sc., F.R.S.

Statistical Department
F. YATES, M.A.
W. G. COCHRAN, M.A.
J. W. WEIL, M.A.

HARPENDEN
HERTS

21st September, 1938.

Professor R. A. Fisher,
Galton Laboratory,
University College,
Gower Street,
London, W. C. 1.

Dear Dr. Fisher,

Some time ago I worked out, for large samples, the efficiency of the test of significance, of a mean of a number of normally distributed values which is obtained by counting only the number of positive and negative signs. This came to $\frac{2}{\pi}$. I have recently been working out the efficiency for samples of any given size and have encountered a point on which I should value your opinion.

The binomial series part of the work seems simple. If $\tau = \mu/\sigma$, the true deviation of the mean relative to its standard error, we have for the binomial series:

$$f = \binom{n}{r} p^r q^{n-r} \qquad \text{where } p = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^\tau e^{-\frac{y^2}{2}}\, dy$$

Hence $$\text{Mean}\left[\frac{\partial^2}{\partial \tau^2} \log f\right] = \frac{-n}{2\pi p q e^{\tau^2}}$$

The question arises: what is the amount of information ~~on~~ in the t-distribution as an estimate of τ? This, it seems to me, should be worked out from the t, τ distribution which you give in the British Association Tables, instead of taking the value which you give at the end of "The Design of Experiments." For a sample of size n

$$\log f = C - \frac{n}{2}\frac{\tau^2}{1+t^2} + \log I_{n-1}\left(-\frac{\tau t\sqrt{n}}{\sqrt{1+t^2}}\right)$$

$$\frac{\partial^2}{\partial\tau^2}\log f = -\frac{n}{1+t^2} - \frac{1}{I_{n-1}^2}\left(\frac{t\sqrt{n}}{\sqrt{1+t^2}}\right)^2 {I'}_{n-1}^{2} + \frac{1}{I_{n-1}}\left(\frac{t\sqrt{n}}{\sqrt{1+t^2}}\right)^2 I''_{n-1}$$

I haven't attempted to average this in general, but the average when $\tau = 0$, which is relevant to the test of significance, can easily be obtained.

We have

$$I_{n-1}(0) = \frac{2^{-\frac{1}{2}(n+1)}}{\left(\frac{n-1}{2}\right)!}, \quad I'_{n-1}(0) = -\frac{2^{-\frac{1}{2}n}}{\left(\frac{n-2}{2}\right)!}, \quad I''_{n-1}(0) = (n-1)\, I_{n-1}(0)$$

The mean value of $\left[-\frac{\partial^2}{\partial\tau^2}\log f\right]$ works out at

$$2\left[\left(\frac{n-1}{2}\right)! \Big/ \left(\frac{n-2}{2}\right)!\right]^2$$

Some values are shown below.

Size of sample	Amount of information
2	1.64
3	2.56
5	4.53
10	9.51
20	19.51
50	49.50
100	99.50

The amount of information looks suspiciously high to me, but I cannot find any mistakes in the working out. What do you think of the method?

With regard to your recent letter about the transformation for ranked data, I will be very interested to see your calculations. I will be in town next Tuesday and will look in at the Galton in case you happen to be free.

Yours sincerely,

W.G. Cochran.