formula will be of use; but it assumes a simple form in the approximation to the cube root of a number R, viz.

$$\frac{a-b}{a+b}=\frac{1}{3}\frac{a^3-R}{a^3+R}.$$

For the n^{th} root of R there is a similar formula

$$\frac{a-b}{a+b}=\frac{1}{n}\frac{a^n-R}{a^n+R}$$

the order of the error in b again being the cube of that in a.

King's College,
Cambridge.

Some Parameters of Sampling Distributions Simply Obtained

By L. M. Brown.

In the theory of statistics a set of quantities $a_1, a_2, \ldots, a_\nu$ is considered, and called a distribution. The moments of this distribution about its origin are defined by the equations

$$\mu_1' = \frac{1}{\nu}\Sigma a_i\ ; \quad \mu_2' = \frac{1}{\nu}\Sigma a_i^2\ ; \quad \mu_3' = \frac{1}{\nu}\Sigma a_i^3.$$

The Mean M of the distribution is defined as μ_1'; if $x_i = a_i - M$, then the moments of the distribution about its mean are defined by the equations

$$\mu_1 = 0\ ; \quad \mu_2 = \frac{1}{\nu}\Sigma x_i^2\ ; \quad \mu_3 = \frac{1}{\nu}\Sigma x_i^3.$$

It is easy to show that $\mu_2 = \mu_2' - M^2$ and that $\mu_3 = \mu_3' - 3M\mu_2' + 2M^3$. The variance, $\sigma^2 = \mu_2$, of the distribution is a measure of its dispersion or spread, and $\beta_1 = \mu_3^2/\mu_2^3$ is a measure of its asymmetry or skewness. All this appears in any elementary account of the subject.

https://doi.org/10.1017/S0950184300000094 Published online by Cambridge University Press

From the above definitions we obtain

$$\mu_2 = \mu_2' - M^2 = \frac{1}{\nu}\Sigma a_i^2 - \frac{1}{\nu^2}(\Sigma a_i)^2$$

$$= \frac{1}{\nu^2}\left[(\nu-1)\,\Sigma a_i^2 - 2\,\Sigma a_i\, a_j\right]$$

$$\mu_3 = \mu_3' - 3M\mu_2' + 2M^3 = \frac{1}{\nu}\,\Sigma a_i^3 - \frac{3}{\nu^2}\,\Sigma a_i\,\Sigma a_i^2 + \frac{2}{\nu^3}\,(\Sigma a_i)^3$$

$$= \frac{1}{\nu^3}\left[(\nu^2 - 3\nu + 2)\,\Sigma a_i^3 - 3\,(\nu - 2)\,\Sigma a_i^2\, a_j + 12\,\Sigma a_i\, a_j\, a_k\right]$$

Let us now take a random sample, say $a_1, a_2, \ldots, a_n$ of n objects of the parent distribution (*the sample is taken without replacement, and so is random but not simple*). The mean of this sample is $b_1 = \frac{1}{n}(a_1 + \ldots + a_n)$. There are ${}^{\nu}C_n$ possible samples, with the same number of possible means; these have a distribution, the sampling distribution of means of n members from the parent distribution. *The object of this note is to obtain by elementary algebra some of the parameters of this distribution, and of the corresponding distribution of variances.*

The mean m_1' of the sampling distribution is given by

$$m_1' = \frac{1}{{}^{\nu}C_n}(b_1 + \ldots) \qquad \text{(summing over the } {}^{\nu}C_n \text{ samples)}$$

where $b_1 = \frac{1}{n}(a_1 + \ldots + a_n)$. The individual a_1 occurs in ${}^{\nu-1}C_{n-1}$ samples, so

$$m_1' = \frac{1}{n\,{}^{\nu}C_n}\,{}^{\nu-1}C_{n-1}\,\Sigma a_i \qquad \text{(summing from 1 to } \nu)$$

$$= \frac{1}{\nu}\Sigma a_i = M. \qquad \text{I}$$

The Mean of the sampling distribution of means is the mean of the parent distribution.

Let m_2', m_3' be the second and third moments of the sampling distribution about the origin, m_2, m_3 the corresponding moments about its mean. Then

$$m_2' = \frac{1}{{}^{\nu}C_n}(b_1^2 + \ldots), \text{ where } b_1^2 = \frac{1}{n^2}\left[(a_1^2 + \ldots + a_n^2) + 2(a_1\, a_2 + \ldots + a_{n-1}\, a_n)\right].$$

Now a term such as a_1^2 comes from a sample containing a_1; there are ${}^{\nu-1}C_{n-1}$ of these. A term such as $a_1 a_2$ comes from a sample containing both a_1 and a_2; there are ${}^{\nu-2}C_{n-2}$ of these. So

$$m_2' = \frac{1}{n^2\, {}^{\nu}C_n}\left[{}^{\nu-1}C_{n-1}\,\Sigma a_i^2 + 2\,{}^{\nu-2}C_{n-2}\,\Sigma a_i a_j\right]$$

$$= \frac{1}{n\nu\,(\nu-1)}\left[(\nu-1)\,\Sigma a_i^2 + 2\,(n-1)\,\Sigma a_i a_j\right].$$

$$M^2 = \frac{1}{\nu^2}\left[\Sigma a_i^2 + 2\,\Sigma a_i a_j\right].$$

$$\therefore\ m_2 = m_2' - M^2 = \frac{\nu - n}{n\nu^2\,(\nu-1)}\left[(\nu-1)\,\Sigma a_i^2 - 2\,\Sigma a_i a_j\right].$$

So the variance of the sampling distribution and the variance of the parent are related by the equation

$$\frac{m_2}{\mu_2} = \frac{\nu - n}{n\,(\nu-1)}. \qquad \text{II}$$

Thus for $1 < n < \nu$, m_2/μ_2 is positive and less than one. *The variance of the sample means is less than the variance of the parent.*

As $\nu \to \infty$, we obtain the well known result $m_2 = \sigma^2/n$.

In the same way, $m_3' = \frac{1}{{}^{\nu}C_n}\,(b_1^3 + \ldots)$, where

$$b_1^3 = \frac{1}{n^3}\left[(a_1^3 + \ldots + a_n^3) + 3\,(a_1^2 a_2 + \ldots) + 6\,(a_1 a_2 a_3 + \ldots)\right].$$ So as before

$$m_3' = \frac{1}{n^3\,{}^{\nu}C_n}\left[{}^{\nu-1}C_{n-1}\,\Sigma a_i^3 + 3\,{}^{\nu-2}C_{n-2}\,\Sigma a_i^2 a_j + 6\,{}^{\nu-3}C_{n-3}\,\Sigma a_i a_j a_k\right]$$

$$= \frac{1}{n^2\nu}\,\Sigma a_i^3 + \frac{3\,(n-1)}{n^2\nu\,(\nu-1)}\,\Sigma a_i^2 a_j + \frac{6\,(n-1)\,(n-2)}{n^2\nu\,(\nu-1)\,(\nu-2)}\,\Sigma a_i a_j a_k.$$

$$M m_2' = \frac{1}{n\nu^2}\,\Sigma a_i^3 + \frac{\nu + 2n - 3}{n\nu^2\,(\nu-1)}\,\Sigma a_i^2 a_j + \frac{6\,(n-1)}{n\nu^2\,(\nu-1)}\,\Sigma a_i a_j a_k.$$

$$M^3 = \frac{1}{\nu^3}\,\Sigma a_i^3 + \frac{3}{\nu^3}\,\Sigma a_i^2 a_j + \frac{6}{\nu^3}\,\Sigma a_i a_j a_k.$$

$$\therefore\ m_3 = m_3' - 3\,M m_2' + 2\,M^3$$

$$= \frac{\nu^2 - 3\,n\nu + 2\,n^2}{n^2\,\nu^3\,(\nu-1)\,(\nu-2)}\left[(\nu-1)\,(\nu-2)\,\Sigma a_i^3 - 3\,(\nu-2)\,\Sigma a_i^2 a_j + 12\,\Sigma a_i a_j a_k\right].$$

$$\therefore\ \frac{m_3}{\mu_3} = \frac{\nu^2 - 3\,n\nu + 2\,n^2}{n^2\,(\nu-1)\,(\nu-2)}. \qquad \text{III}$$

So the parameter β_1 of the parent distribution and the corresponding parameter b_1 of the distribution of sample means are related by the equation

$$\frac{b_1}{\beta_1} = \frac{m_3^2}{m_2^3} \cdot \frac{\mu}{\mu_3^2} = \frac{1}{n} \cdot \frac{(\nu-1)(\nu-2n)^2}{(\nu-n)(\nu-2)^2}. \qquad \text{IV}$$

Thus for $1 < n < \nu - 1$, b_1/β_1 is less than 1. *The skewness of the sample means is less than the skewness of the parent*; in fact, as n grows from 1 to $\frac{1}{2}\nu$ it decreases steadily from 1 to 0. The sign of m_3 is the same as the sign of μ_3 if $n < \frac{1}{2}\nu$, but is opposite if $n > \frac{1}{2}\nu$. If $\nu \to \infty$, then $b_1/\beta_1 \to 1/n$.

Consider now the *variances* of the samples. The variance s_1^2 of the sample $a_1, a_2, \ldots, a_n$ is

$$s_1^2 = \frac{1}{n}(a_1^2 + \ldots + a_n^2) - \frac{1}{n^2}(a_1 + \ldots + a_n)^2$$
$$= \frac{1}{n^2}\Big[(n-1)(a_1^2 + \ldots + a_n^2) - 2(a_1 a_2 + \ldots + a_{n-1} a_n)\Big].$$

Let us calculate the mean M_{s^2} of these variances of the ${}^\nu C_n$ samples.

$$M_{s^2} = \frac{1}{{}^\nu C_n \, n^2}\Big[{}^{\nu-1}C_{n-1}(n-1)\Sigma a_i^2 - {}^{\nu-2}C_{n-2}\, 2\, \Sigma a_i a_j\Big]$$
$$= \frac{n-1}{n\nu(\nu-1)}\Big[(\nu-1)\Sigma a_i^2 - 2\Sigma a_i a_j\Big].$$

But the variance of the parent distribution is given by

$$\sigma^2 = \frac{1}{\nu^2}\Big[(\nu-1)\Sigma a_i^2 - 2\Sigma a_i a_j\Big].$$

So the mean of the sample variances is related to the variance of the parent distribution by the equation

$$\frac{M_{s^2}}{\sigma^2} = \frac{n-1}{n}\frac{\nu}{\nu-1}. \qquad \text{V}$$

The mean of the sample variances is always less than the variance of the parent; in fact, if we denote the variance m_2 of the distribution of means by σ_m^2, equations II and V lead to the result

$$\sigma^2 - M_{s^2} = \sigma_m^2. \qquad \text{VI}$$

As $\nu \to \infty$, equation V reduces to the well known result $M_{s^2} = (n-1)\sigma^2/n$.

THE ROYAL TECHNICAL COLLEGE,
GLASGOW.