STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 408
SP 112

Percentiles and approximations of the sample median over the sample range for samples of size 3,5,7, and 9 from a standard normal distribution

N. Bouma and A. Vehmeyer

Reprinted from

Statistica Neerlandica, 23(1969), p 235-239

1969

Percentiles and approximations of the sample median over the sample range for samples of size 3, 5, 7 and 9 from a standard normal distribution

by N. BOUMA and A. VEHMEYER *

Summary A table of the distribution function of the quotient of sample median and sample range of small samples from the normal distribution is given which is based partly on theoretical results published by L. DE HAAN and J. TH. RUNNENBURG, partly on numerical results obtained with Monte Carlo methods. Moreover an approximation of this distribution by Students distributions is introduced.

1. Introduction

In this paper the attention is restricted to samples of size 3, 5, 7 or 9 from the normal distribution with $\mu = 0, \sigma = 1$. The purpose is to give numerical results about the random variable, the quotient of sample median and sample range, mainly to enable it to be used as an easy-to-calculate statistic. The authors made use of the theoretical results DE HAAN and RUNNENBURG obtained in [1]. A great deal of the computational work was programmed in Algol 60 and done by the Electrologica X8 computer of the Mathematical Centre.

The following notations are introduced:

n	an odd natural number
$\underline{x}_{(1)}, \ldots, \underline{x}_{(n)}$	an ordered random sample of size n from a standard normal distribution
$\underline{m}_n = \underline{x}_{(n/2+1/2)}$	the sample median
$\underline{r}_n = \underline{x}_{(n)} - \underline{x}_{(1)}$	the sample range
$\underline{h}_n = \underline{m}_n / \underline{r}_n$	the investigated random variable
f_n	the probability density of $\underline{h}_n$
F_n	the distribution function of $\underline{h}_n$

Furthermore, let the symbols $\hat{=}$, $\hat{\approx}$ stand for „has the same distribution as", „has approximately the same distribution as" resp.. The random variables $\underline{u}$, $\underline{t}_\nu$, $\underline{\chi}_\nu^2$ are supposed to have respectively the normal ($\mu = 0, \sigma = 1$), Student's and the chi-square distribution with ν degrees of freedom.

2. The distribution function of $\underline{h}_n$

In their paper on the quotient of sample median and sample range DE HAAN and

Report S 408 (SP 112), Statistical Department, Mathematical Centre, Amsterdam.
* Mathematical Centre, Amsterdam.

RUNNENBURG derived formula's for the distribution function $F_3(t)$ ([1] formula 11)

$$F_3(t) = \frac{1}{2} - \frac{3}{2\pi}\left\{\arccos\left(\frac{t-1}{\sqrt{4t^2+2}}\right)+\arccos\left(\frac{t+1}{\sqrt{4t^2+2}}\right)\right\} \tag{1}$$

and for the density $f_5(t)$ ([1] formula 15).

From this density, values of the distribution function $F_5(t)$ can easily be computed by means of numerical integration. For other values of n theoretical results could not be obtained. Therefore the method of simulating observations was used for $n = 7$ and $n = 9$. For $n = 3, 5, 7, 9$ a large number (at least 30,000) of observations of $\underline{h}_n$ was generated and grouped in 10,000 classes. After that the cumulative frequency table was constructed and used as a numerical approximation of the real distribution function $F_n(t)$ of $\underline{h}_n$.

In using simulation the main problem is to get an idea of the accuracy of the obtained results. In this case there were two indications that the maximum error is less than .01.

Table 1 Percentiles of the distribution function $F_n(t)$ of the sample median over the sample range for samples of size n from the standard normal distribution

	n					n			
$F_n(t)$	3	5	7	9	$F_n(t)$	3	5	7	9
.50	.000	.000	.00	.00	.75	.292	.162	.12	.09
.51	.010	.006	.00	.00	.76	.306	.170	.12	.10
.52	.020	.012	.01	.01	.77	.322	.179	.13	.10
.53	.031	.018	.01	.01	.78	.338	.187	.14	.11
.54	.042	.024	.02	.01	.79	.354	.196	.14	.11
.55	.053	.030	.02	.02	.80	.372	.205	.15	.12
.56	.063	.036	.03	.02	.81	.390	.214	.15	.12
.57	.074	.042	.03	.02	.82	.409	.224	.16	.13
.58	.085	.048	.03	.03	.83	.430	.234	.17	.14
.59	.095	.054	.04	.03	.84	.451	.245	.18	.14
.60	.106	.060	.04	.03	.85	.475	.256	.18	.15
.61	.117	.066	.05	.04	.86	.500	.268	.19	.15
.62	.128	.073	.05	.04	.87	.527	.281	.20	.16
.63	.140	.079	.06	.05	.88	.556	.295	.21	.17
.64	.151	.085	.06	.05	.89	.589	.309	.22	.18
.65	.163	.092	.07	.05	.90	.626	.325	.23	.19
.66	.174	.098	.07	.06	.91	.668	.343	.25	.20
.67	.186	.105	.08	.06	.92	.717	.363	.26	.21
.68	.198	.112	.08	.07	.93	.775	.385	.27	.22
.69	.211	.118	.09	.07	.94	.844	.410	.29	.23
.70	.223	.125	.09	.07	.95	.932	.441	.31	.25
.71	.236	.132	.10	.08	.96	1.05	.478	.33	.26
.72	.249	.140	.10	.08	.97	1.22	.528	.36	.29
.73	.263	.147	.11	.09	.98	1.50	.600	.41	.32
.74	.277	.155	.11	.09	.99	2.14	.735	.49	.37
.975	1.34	.56	.37	.30	.995	3.03	.89	.58	.44

1. The real distribution function $F_n(t)$ of $\underline{h}_n$ is symmetric.
   So the cumulative frequency function, too, had to show this symmetry. This condition was fulfilled with rather good accuracy.
2. For $n = 3$ and $n = 5$, the results obtained by simulating could be compared with the computed results of which the precision was known.
   The maximum error was smaller than .01.

Thus, table 1 contains the computed values of the percentiles of $F_n(t)$ for $n = 3$ and $n = 5$, and the equivalent results obtained by means of simulation for $n = 7$ and $n = 9$. Because of the symmetry of the distribution functions only the upper fifty percentiles are given.

3. The moments

As was shown by DE HAAN and RUNNENBURG [1] only the first $n-2$ moments of $\underline{h}_n$ exist. The existing second and fourth moments, calculated with numerical integration methods, are given in tabel 2.

Table 2 The existing second and fourth moments of $\underline{h}_n$

	$n = 3$	$n = 5$	$n = 7$	$n = 9$
μ_2	–	.0824	.0373	.0224
μ_4	–	–	.0069	.0019

4. Approximation by Student's distribution

In [2] PEARSON mentions that the square of the sample range $\underline{r}_n^2$ can be approximated by a chi-square distribution. If α_n and ν_n are chosen in such a way that for $i = 1,2$ the i-th moment of $\underline{r}_n^2$ is equal to the i-th moment of the random variable $\alpha_n \underline{\chi}_{\nu_n}^2$, or equivalently, if

$$\alpha_n = \frac{\sigma^2(\underline{r}_n^2)}{2\mathscr{E}\underline{r}_n^2} \quad \text{and} \quad \nu_n = \frac{2(\mathscr{E}\underline{r}_n^2)^2}{\sigma^2(\underline{r}_n^2)}, \tag{2}$$

then the distribution function of $\alpha_n \underline{\chi}_{\nu_n}^2$ is a good approximation of the distribution of $\underline{r}_n^2$. Thus of the range itself it can be said:

$$\underline{r}_n \mathrel{\underset{\sim}{\approx}} \sqrt{\alpha_n} \underline{\chi}_{\nu_n}. \tag{3}$$

Besides

$$\frac{\underline{m}_n}{\sigma(\underline{m}_n)} \mathrel{\underset{\sim}{\approx}} \underline{u} \tag{4}$$

and $\underline{m}_n$ and $\underline{r}_n$ have correlation coefficient 0; thus one may expect,

$$\underline{h}_n = \frac{\underline{m}_n}{\underline{r}_n} \mathrel{\underset{\sim}{\approx}} \frac{\sigma(\underline{m}_n)}{\sqrt{\alpha_n}} \frac{\underline{u}}{\underline{\chi}_{\nu_n}} \mathrel{\underset{\sim}{\cong}} \frac{\sigma(\underline{m}_n)}{\sqrt{\alpha_n \nu_n}} \underline{t}_{\nu_n}$$

Or

$$C_n \underline{h}_n \stackrel{\approx}{\sim} \underline{t}_{\nu_n} \tag{5}$$

with

$$C_n = \frac{\mathscr{E}\underline{r}_n^2}{\sigma(\underline{m}_n)} \quad \text{and} \quad \nu_n = \frac{2(\mathscr{E}\underline{r}_n^2)^2}{\sigma^2(\underline{r}_n^2)} \tag{6}$$

The moments of $\underline{r}_n$ and $\underline{m}_n$ can be found in PEARSON and HEARTLEY [3] and SARHAN and GREENBERG [4].

However this approximation of $\underline{h}_n$ can still be improved by choosing C_n and ν_n anew in order to equalize some of the moments of $C_n\underline{h}_n$ and $\underline{t}_{\nu_n}$. Of course this can only be done for the existing moments of $\underline{h}_n$. The numerical values of C_n and ν_n changed very little.

The authors' aim however was to obtain a simple-to-use approximation. So the attention was restricted to Student distributions with an integer number of degrees of freedom. Thus for ν_n an integer near to the values computed as mentioned above was taken and, in view of the non-existance of some of the moments, C_n was chosen in such a way that for $F_n(t) = .05$ the values of the obtained approximation coincides with the distribution function of $\underline{h}_n$. Whenever there had to be chosen between two values of ν_n the errors at the points $F_n(t) = .1$, $F_n(t) = .025$, $F_n(t) = .01$ were used in order to decide.

Table 3 gives the values of ν_n and C_n according to (6), improved by means of equalizing moments and, finally, with integer ν_n. The last columns can be used without interpolation for ν in a table of Student's distribution.

Table 3 Approximation by Student's distribution; $C_n\, \underline{h}_n$ is approximated by $\underline{t}_{\nu_n}$

	C_n and ν_n computed according to (6)		improved approximation		approximation with integer ν_n	
n	ν_n	C_n	ν_n	C_n	ν_n	C_n
3	2.0	2.854	2.38	2.71	2	3.13
5	3.8	4.634	4.610	4.402	5	4.56
7	5.5	6.169	6.777	5.963	7	6.1
9	7.0	7.553	9.178	7.343	9	7.3

5. Final remark

Another method for computing the values of the distribution function $F_n(t)$ of the quotient of sample median and sample range is still being tried out. Let ψ, Φ denote the density and distribution function of the normal distribution ($\mu = 0$, $\sigma = 1$), then the marginal density of $\underline{x}_{(1)}$, $\underline{m}_n$, $\underline{x}_{(n)}$ can be written as (see e.g. [4])

$$g(x_{(1)}, m_n, x_{(n)}) = \frac{n!}{\left[\left(\frac{n-1}{2}\right)!\right]^2} [(\Phi(x_{(n)}) - \Phi(m_n))(\Phi(m_n) - \\ - \Phi(x_{(1)})]^{(n-1)/2} \psi(x_{(1)})\psi(m_n)\psi(x_{(n)}) \quad \text{for} \quad x_{(1)} < m_n < x_{(n)} \quad 0 \text{ else.}$$

Because

$$F(t) = P\{\underline{h}_n \leqslant t\} = P\left\{\frac{\underline{m}_n}{\underline{r}_n} \leqslant t\right\} = P\{\underline{m}_n \leqslant t(x_{(n)} - x_{(1)})\}$$

the values of $F_n(t)$ could be calculated by means of numerical integration of the density $g(x_{(1)}, m_n, x_{(n)})$ over the set

$$\{(x_{(1)}, m_n, x_{(n)}) \mid m_n \leqslant t(x_{(n)} - x_{(1)})\}.$$

The usual procedures for numerical integration to our disposal however do not yet yield satisfactory results for this complicated multiple integral.

6. Acknowledgements

The investigation of the quotient of sample median and sample range was started and supervised by Professor HEMELRIJK, whom the authors wish to thank for his stimulating guidance.

References

[1] L. DE HAAN and J. TH. RUNNENBURG, Some remarks concerning the quotient of sample median and sample range for a sample of size $2n+1$ from a normal distribution. Statistica Neerlandica **23** (1969) 227–234.

[2] E. S. PEARSON, The probability integral of the range in samples of n observations from a normal population. Biometrika **32** (1942) 301–310.

[3] E. S. PEARSON and H. O. HARTLEY, Biometrika tables for statisticians. Cambrigde University Press (1958).

[4] A. E. SARHAN and B. G. GREENBERG, Contributions to order statistics. John Wiley and Sons (1962).