

# On the Mean Difference Variance in Random Samples of Student's Variables

Fabio Manca, Claudia Marin

Department of Education, Psychology, Communication, University of Bari Aldo Moro, Bari, Italy

Email: [claudia.marin@uniba.it](mailto:claudia.marin@uniba.it)

**How to cite this paper:** Manca, F. and Marin, C. (2020) On the Mean Difference Variance in Random Samples of Student's Variables. *Open Journal of Statistics*, 10, 659-663.

<https://doi.org/10.4236/ojs.2020.104040>

**Received:** July 9, 2020

**Accepted:** August 7, 2020

**Published:** August 10, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

The purpose of this paper is to obtain the expression of the sample mean difference variance of the Student's distributive model. In the 2007 the study of the mean difference variance, after some decades, was resumed by Campobasso [1]. Using the Nair's [2] and Lomnicki's general results [3], he obtained the variance of sample mean difference for different distributive models (Laplace's, triangular, power, logit, Pareto's and Gumbel's model). In addition he extended the knowledge comparing to the ones already known for the other distributive model (normal, rectangular and exponential model).

## Keywords

Mean Difference Variance, Random Sample, Student

## 1. The Mean Difference Variance General Expression

Let  $X$  be a continuous random variable with density function  $f(x)$  and distribution function  $F(x)$ . Then, let  $X_1, X_2, \dots, X_n$  be a simple random sample from such population; the sample mean difference is

$$\bar{\Delta} = \frac{\sum_{i=1}^n \sum_{j=1}^n |X_i - X_j|}{n(n-1)} \quad (1)$$

The mean value of  $\bar{\Delta}$  is equal to the mean difference of the population

$$\Delta = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x - y| f(x) f(y) dx dy \quad (2)$$

In 1952 Lomnicki [3] obtained the following general expression of the sample mean difference variance

$$\text{var}(\bar{\Delta}) = \frac{1}{n(n-1)} [4(n-1)\sigma^2 + 16(n-2)I - 2(2n-3)\Delta^2] \tag{3}$$

in which  $\sigma^2$  and  $\Delta$  are the variance and the mean difference of the considered distributive model respectively, whereas

$$I = \int_{-\infty}^{+\infty} G(x)H(x)f(x)dx \tag{4}$$

in which

$$G(x) = \int_{-\infty}^x (x-y)f(y)dy \tag{5}$$

and

$$H(x) = G(x) + \mu - x, \tag{6}$$

in which  $\mu$  is the mean value of such distributive model.

The mean value  $\mu$  and the variance are known for almost all the distributive models. Concerning the mean difference  $\Delta$ , the known results are collected in Girone's and Mazzitelli's paper [4].

So, to determine the expression of the sample mean difference variance it's only needed the calculation of  $I$ .

## 2. I Expression for Odd $g$

The density function of the Student's distributive model is

$$f(x) = \frac{(1+x^2/2)^{\frac{g+1}{2}}}{\sqrt{g}B(g/2, g/2)}, \quad -\infty < x < +\infty \tag{7}$$

in which the parameter  $g$  is called number of degrees of freedom.

Using *Mathematica* software for such model for  $g = 5, 7, \dots, 19$  the obtained values of  $I_g$  are shown in **Table 1**.

The second term is easily represented by the following formula

**Table 1.** The obtained values of  $I_g$  for  $g = 5, 7, \dots, 19$ .

$g$	values di $I_g$
3	$15/(2\pi^2) - 1/2$
5	$1925/(432\pi^2) - 5/18$
7	$17,017/(4500\pi^2) - 7/30$
9	$9,561,123/(2,744,000\pi^2) - 3/14$
11	$2,369,851/(714,420\pi^2) - 11/54$
13	$2,170,568,075/(67,6190,592\pi^2) - 13/66$
15	$3,920,876,125/(1,250,497,248\pi^2) - 5/26$
17	$1,077,676,328,213/(349,825,132,800\pi^2) - 17/90$
19	$135,999,445,173,949/(44,757,574,933,500\pi^2) - 19/102$

$$-\frac{g}{6(g-2)}. \tag{8}$$

The first term expression is more complicated to be determined. After several attempts comparing each first term to the previous one, we pointed out the recurring formula:

$$A_g = A_{g-2} \frac{g(g-4)(g-3)^2(3g-4)(3g-8)}{(g-2)^4(3g-7)(3g-11)}, \text{ per } g = 5, 7, \dots \tag{9}$$

with the initial value  $A_3 = 15/(2\pi^2)$ .

Considering the previous relation, we came to the following expression of the first  $I_g$  term for odd  $g$  values greater than 3:

$$A_g = \frac{\sqrt{3}g\Gamma[(g-1)/2]^2 \Gamma(g/2-1/3)\Gamma(g/2+1/3)}{2(g-2)\Gamma(g/2)^2 \Gamma(g/2-1/6)\Gamma(g/2-5/6)\pi}, \tag{10}$$

and then the  $I_g$  expression

$$I_g = \frac{\sqrt{3}g\Gamma[(g-1)/2]^2 \Gamma(g/2-1/3)\Gamma(g/2+1/3)}{2(g-2)\Gamma(g/2)^2 \Gamma(g/2-1/6)\Gamma(g/2-5/6)\pi} - \frac{g}{6(g-2)} \tag{11}$$

### 3. I Expression for Even g

Using again *Mathematica* software for  $g = 4, 6, \dots, 20$  the obtained values of  $I_g$  are shown in **Table 2**.

The second term of  $I_g$  is represented again by the simple formula

$$-\frac{g}{6(g-2)}. \tag{12}$$

After several attempts, comparing each first term to the previous one, we pointed out the recurring formula:

$$A_g = A_{g-2} \frac{g(g-4)(g-3)^2(3g-4)(3g-8)}{(g-2)^4(3g-7)(3g-11)}, \text{ per } g = 6, 8, \dots \tag{13}$$

**Table 2.** The obtained values of  $I_g$  for  $g = 4, 6, \dots, 20$ .

$g$	values di $I_g$
4	8 / 15 - 1/3
6	9/22 - 1/4
8	8000/21,879 - 2/3
10	30,625/89,148 - 5/24
12	1,778,112/5,386,025 - 1/5
14	166,012/516,925 - 7/36
16	345,506,304/1,097,845,315 - 1/6
18	6,832,522,125/22,049,643,544 - 3/16
20	8,450,958,230,000/27,608,909,922,531 - 5/27

with the initial value  $A_4 = 8/15$ . It has to be noticed that the recurring formula is the same one as the odd case.

Considering the previous relation, we came to the following expression of the first  $I_g$  term for even  $g$  values greater than 4:

$$A_g = \frac{2^{3-2g} \sqrt{3} g (g-2) \Gamma(g-2)^2 \Gamma(g/2-1/3) \Gamma(g/2+1/3)}{\Gamma(g/2)^4 \Gamma(g/2-1/6) \Gamma(g/2-5/6)} \tag{14}$$

and then the  $I_g$  expression

$$I_g = \frac{2^{3-2g} \sqrt{3} g (g-2) \Gamma(g-2)^2 \Gamma(g/2-1/3) \Gamma(g/2+1/3)}{\Gamma(g/2)^4 \Gamma(g/2-1/6) \Gamma(g/2-5/6)} - \frac{g}{6(g-2)} \tag{15}$$

Through some algebraic steps it is easily verified that the two  $I_g$  formulas for the odd case and the even one are the same and, moreover, a single more compact expression is the following

$$I_g = \frac{2^{3-2g} \sqrt{3} g \Gamma(g/2-1/3) \Gamma(g/2+1/3)}{(g-1)^2 (g-2) B(g/2, g/2) \Gamma(g/2-1/6) \Gamma(g/2-5/6)} - \frac{g}{6(g-2)} \tag{16}$$

#### 4. The Sample Mean Difference Variance

Let us remind that for the Student's distributive model the expressions of the mean value ( $\mu$ ), the variance ( $\sigma^2$ ) and the mean difference ( $\Delta$ ) are the following:

$$\mu = 0, \tag{17}$$

$$\sigma^2 = \frac{g}{g-2}, \tag{18}$$

$$\Delta = \frac{\sqrt{g} B[(g-1)/2, (g+1)/2]}{(2g-1) B(g, g) B(g/2, g/2) 2^{2g-3}}. \tag{19}$$

Using the Lomnicki's formula we came to the following expression of the mean difference variance for the Student's distributive model:

$$\begin{aligned} \text{var}(\bar{\Delta}) &= \frac{4g(n+1)}{3(g-2)n(n-1)} + \frac{2^{5-2g} g \Gamma[g-1/2]^2 \Gamma[(g-1)/2]^2 (3-2n)}{\Gamma[g/2]^6 n(n-1)} \\ &+ \frac{8\sqrt{3} g \Gamma[(g-1)/2]^2 h(g)(n-2)}{\pi(g-2) \Gamma[g/2]^2 n(n-1)} \end{aligned} \tag{20}$$

in which

$$h(g) = \frac{\Gamma(g/2-1/3) \Gamma(g/2+1/3)}{\Gamma(g/2-1/6) \Gamma(g/2-5/6)} \tag{21}$$

It is easily checked that, as  $g$  diverges,

$$\text{var}(\bar{\Delta}) = \frac{72 - 48\sqrt{3} + 4\pi - (48 + 24\sqrt{3} + 4\pi)n}{3\pi n(n-1)}, \tag{22}$$

that represents the sample difference variance for the normal model. It is also easily verified that, as  $n$  diverges, the above mentioned variance approaches zero, which means that  $\bar{\Delta}$  is also a consistent estimator.

## 5. Conclusions

The sample mean difference  $\bar{\Delta}$  is a correct estimator of the mean difference population for every distributive model. To verify if it is also consistent or not we need to calculate its variance, in this paper we have obtained the variance of  $\bar{\Delta}$  formal expression for the Student's distributive model in terms of the parameter  $g$  (degrees of freedom) and of the sample size  $n$ .

Because, even for the Student's distributive model, such variance approaches zero as the sample size  $n$  diverges,  $\bar{\Delta}$  results consistent. As  $g$  diverges the Student's distributive model tends to the normal one. As a matter of fact the variance of  $\bar{\Delta}$  expression we found approaches the variance of  $\bar{\Delta}$  for the normal distributive model.

## Acknowledgements

The helpful and constructive comments of a referee which lead to an improvement of the presentation of the paper and support from the editorial staff of Open Journal of Statistics to process the paper are all gratefully acknowledged.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Campobasso, F. (2007) Alcuni risultati sulla varianza della differenza media campionaria. *Annals of the Department of Statistical Sciences*, **6**.
- [2] Nair, U.S. (1936) The Standard Error of Gini's Mean Difference. *Biometrika*, **28**, 428-436. <https://doi.org/10.2307/2333957>
- [3] Lomnicki, Z.A. (1952) The Standard Error of Gini's Mean Difference. *Annals of Mathematical Statistics*, **23**, 635-637. <https://doi.org/10.1214/aoms/1177729346>
- [4] Girone, G. and Mazzitelli, D. (2007) La differenza media nei principali modelli distributivi continui. *Annals of the Department of Statistical Sciences*, **6**.