SCIENTIFIC PAPERS
OF THE UNIVERSITY OF PARDUBICE
Series A
Faculty of Chemical Technology
5 (1999)

APPLICATION OF QUANTILES AND PIVOTS IN EVALUATION OF SMALL SAMPLE SIZES OF DATA

Josef DEMPÍR
CZ-285 32 Hlízov 22

Received March 15, 1999

In the present paper the use of quantiles and pivots in evaluation of small sample sizes of data is decribed. Formulae for computing mean value and standard deviation are derived. Some examples are given to illustrate the usefulness of this approach in the analytical practice.

Introduction

In evaluation of small sample sizes of analytical results it may often happen that the estimates of the mean and the standard deviation are biased. This is usually caused by the influence of some extreme values of the results, especially if these results cannot be identified by statistical tests as outliers. This can be solved by the use of robust statistics (cf. Ref. [1]), as it is to some extent insensitive to the presence of the lowest and/or highest values. The use of quartiles [2] and pivots [3] represents a simple solution to such problem. The aim of the present paper is to provide a useful guide, in the form of tables, for actual application of quartiles and pivots and, in addition to that, to illustrate the applicability of this approach by some typical examples from analytical practice.

The Use of Quartiles

For a set of data arranged in ascending order the position p of the lower quartile $\tilde{x}_{0.25}$ and that of q for an upper quartile $\tilde{x}_{0.75}$ are defined by the following relationships

$$p = 0.5 + 0.25n \tag{1}$$

$$q = 0.5 + 0.75n \tag{2}$$

where n is the total number of the data [2]. The positions of the two quartiles, calculated by means of a linear interpolation, are given in Table I.

Table I Positions of quartiles and pivots for n = 4 to 12

n	p	q	$\tilde{x}_{0.25}$	$\tilde{x}_{0.75}$	Pi_{lo}	Pi_{up}
4	1.50	3.50	$0.50x_1 + 0.50x_2$	$0.50x_3 + 0.50x_4$	x_1	x_4
5	1.75	4.25	$0.25x_1 + 0.75x_2$	$0.750x_4 + 0.25x_5$	x_2	x_4
6	2.00	5.00	$1.00x_2$	$1.00x_5$	x_2	x_5
7	2.25	5.75	$0.75x_2 + 0.25x_3$	$0.25x_5 + 0.75x_6$	x_2	x_6
8	2.50	6.50	$0.50x_2 + 0.50x_3$	$0.50x_6 + 0.50x_7$	x_2	x_7
9	2.75	7.25	$0.25x_2 + 0.75x_3$	$0.75x_7 + 0.25x_8$	x_3	x_7
10	3.00	8.00	$1.00x_3$	$1.00x_8$	x_3	x_8
11	3.25	8.75	$0.75x_3 + 0.25x_4$	$0.25x_8 + 0.75x_9$	x_3	x_9
12	3.50	9.50	$0.50x_3 + 0.50x_4$	$0.50x_9 + 0.50x_{10}$	x_3	x_{10}

The interquartile range, i.e. $\tilde{x}_{0.75} - \tilde{x}_{0.25}$, can be used to estimate a mean value $\bar{x}$ and a standard deviation s as follows

$$\bar{x} = \frac{1}{2}\left(\tilde{x}_{0.75} + \tilde{x}_{0.25}\right) \tag{3}$$

where $z_{0.75}$ is a quantile of the standardized normal distribution taken from the probability $P = 0.75$ and $z_{0.25}$ represents an analogous quantile for $P = 0.25$. As can be found in statistical tables (e.g. Ref. [4]), $z_{0.75} = 0.67449$ and $z_{0.25} = -0.67449$. Inserting these values into Eq. (4) yields

$$s = \frac{1}{z_{0.75} - z_{0.25}}\left(\tilde{x}_{0.75} - \tilde{x}_{0.25}\right) \tag{4}$$

$$s = 0.7413\left(\tilde{x}_{0.75} - \tilde{x}_{0.25}\right) \tag{5}$$

The expressions for $\bar{x}$ and s varying from 4 to 12 are listed in Table II.

Table II Estimates of the mean values and the standard deviations using the quartiles for $n = 4$ to 12

n	Mean values	Standard deviations
4	$0.50[0.50(x_4 + x_1) + 0.50(x_3 + x_2)]$	$0.7413[0.50(x_4 - x_1) + 0.50(x_3 - x_2)]$
5	$0.50[0.25(x_5 + x_1) + 0.75(x_4 + x_2)]$	$0.7413[0.25(x_5 - x_1) + 0.75(x_4 - x_2)]$
6	$0.50(x_5 + x_2)$	$0.7413(x_5 - x_2)$
7	$0.50[0.75(x_6 + x_2) + 0.25(x_5 + x_3)]$	$0.7413[0.75(x_6 - x_2) + 0.25(x_5 - x_3)]$
8	$0.50[0.50(x_7 + x_2) + 0.50(x_6 + x_3)]$	$0.7413[0.50(x_7 - x_2) + 0.50(x_6 - x_3)]$
9	$0.50[0.25(x_8 + x_2) + 0.75(x_7 + x_3)]$	$0.7413[0.25(x_8 - x_2) + 0.75(x_7 - x_3)]$
10	$0.50(x_8 + x_3)$	$0.7413(x_8 - x_3)$
11	$0.50[0.75(x_9 + x_3) + 0.25(x_8 + x_4)]$	$0.7413[0.75(x_9 - x_3) + 0.25(x_8 - x_4)]$
12	$0.50[0.50(x_{10} + x_3) + 0.50(x_9 + x_4)]$	$0.7413[0.50(x_{10} - x_3) + 0.50(x_9 - x_4)]$

The Use of Pivots

The concept of pivots was introduced in the statistical procedures by Horn [3]. For the position p of the lower pivot Pi_{lo} two relations are given (6) and (7). For the choice between them it is decisive whether (6) or (7) leads to an integer number of p [5].

$$p = \begin{cases} \{\text{int}[(n+1)/2]\}/2 & \text{or} \\ \{\text{int}\ [(n+1)/2] + 1\}/2 \end{cases} \tag{6), (7}$$

The position q of the upper pivot Pi_{up} is

$$q = n + 1 - p \tag{8}$$

$$p + q = n + 1 \tag{9}$$

Experimental data corresponding to the lower and upper pivots can be found in Table I.

The interpivot range ($Pi_{up} - Pi_{lo}$) can be used — similarly as the interquartile range — for robust estimates of the mean value and standard deviation of a set of data.

The mean value $\bar{x}$ is given by a simple relation

$$\bar{x} = 0.5\left(Pi_{up} + Pi_{lo}\right) \tag{10}$$

How to proceed in order to compute standard deviations using pivots is shown in the following example. For $n = 4$ the lower pivot is x_1 and the upper pivot is x_4, from which it follows for $p = 1$ and $q = 4$. The probability for x_1 denoted by P_1 and that for x_4 denoted by P_4 are given as

$$P_1 = \frac{p - 0.5}{n} = \frac{1 - 0.5}{4} = 0.125 \tag{11}$$

$$P_4 = \frac{q - 0.5}{n} = \frac{4 - 0.5}{4} = 0.875 \tag{12}$$

For the standard deviation s using interpivot range we have

$$s = \frac{1}{2\,|z_p|}\left(Pi_{up} - Pi_{lo}\right) \tag{13}$$

The corresponding values of z_p are respectively [4]: $z_{0.125} = -1.150349$ and $z_{0.875} = 1.150349$. For our example $n = 4$ we obtain

$$s = \frac{x_4 - x_1}{2} \times 1.150349 = 0.4347\left(x_4 - x_1\right) \tag{14}$$

The resulting formulae for computation of robust estimates of $\bar{x}$ and s using interpivot are given in Table III.

Table III Estimates of the mean values and the standard deviations using pivots for $n = 4$ to 12

n	Mean values	Standard deviations
4	$0.50(x_4 + x_1)$	$0.4347(x_4 - x_1)$
5	$0.50(x_4 + x_2)$	$0.9535(x_4 - x_2)$
6	$0.50(x_5 + x_2)$	$0.7413(x_5 - x_2)$
7	$0.50(x_6 + x_2)$	$0.6316(x_6 - x_2)$
8	$0.50(x_7 + x_2)$	$0.5636(x_7 - x_2)$
9	$0.50(x_7 + x_3)$	$0.8483(x_7 - x_3)$
10	$0.50(x_8 + x_3)$	$0.7413(x_8 - x_3)$
11	$0.50(x_9 + x_3)$	$0.6686(x_9 - x_3)$
12	$0.50(x_{10} + x_3)$	$0.6155(x_{10} - x_3)$

Examples

The use of the above-described robust methods and their comparison with classical evaluation procedure is shown on practical examples. The experimental data were taken from the paper by Govindaraju and cowokers [6]. Trace amounts of Cd, Ba, Er, Nb, As, Pb, Ce, La and Mo were chosen as examples. The analytical methods used have the following codes: (see Table IV) : AA – atomic absorption spectroscopy, NM – nuclear methods, mostly neutron activation analysis, SM – mass spectrometry.

In Table V statistical parameters of the experimental data are derived using three procedures: the classical one and the interquartile and interpivot ranges.

Table IV Examples. Data for trace elements

Element	Method	Experimental results, ppm				
Cd	AA	0.31	1.00	1.00	4.00	
Ba	AA	45.00	49.00	50.00	54.00	65.00
Er	SP	6.40	6.60	7.21	7.55	7.60
		7.73				
Nb	SP	125	151	170	185.25	193
		208	257.09			
As	NM	28.15	30.00	30.00	32.00	32.00
		33.00	34.36	52.67		
Pb	AA	63.00	67.08	75.00	80.00	80.00
		80.00	87.50	95.00	116.00	
Ce	NM	72.00	75.00	84.42	89.22	96.50
		97.45	104.40	106.00	106.00	109.00
La	NM	24.25	25.45	26.35	27.34	28.00
		28.00	29.30	29.70	29.90	30.00
		33.00				
Mo	SM	3.50	3.60	3.92	4.00	4.00
		4.20	4.46	4.50	4.60	4.60
		4.70	5.80			

Table V Examples. Evaluation of trace elements — ppm

Element	Standard procedure		Quartiles		Pivots	
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
Cd	1.58	1.65	1.58	1.37	2.16	1.60
Ba	52.60	7.64	52.38	6.49	51.50	4.77
Er	7.18	0.56	7.10	0.74	7.10	0.74
Nb	184	42	180.00	35.95	179.5	36.00
As	34.02	7.78	31.84	2.73	32.18	2.46
Pb	82.62	15.80	81.20	12.12	81.25	10.60
Ce	94.0	13.33	95.2	16.0	95.2	16.0
La	28.3	2.45	28.2	2.41	28.13	2.37
Mo	4.32	0.61	4.28	0.47	4.26	0.42

Conclusion

For correct evaluation of experimental data the normality of the results, i.e. the absence of outliers and significant skewness are of primordial importance. If the normality of data is fulfilled, statistical methods used do not play important role: both classical approach and robust procedures yield practically identical values of the mean and standard deviation.

Complications are met in cases when the normality of data cannot be accepted. Robust methods using interquartile and interpivot range can be useful providing that the formulae used for computing the mean value and the standard deviation do NOT contain the minimum and/or maximum results. Using the interquartile range this is true for $n \geq 5$, with interpivot range for $n \geq 6$.

Acknowledgements

Professor Stanislav Kotrlý, University of Pardubice, is thanked for his helpful comments and critical reviewing the manuscript.

References

1. Huber P.J..: *Robust Statistics*, Willey, New York 1981.
2. Danzer K.: Fresenius Zeit. Anal. Chem. **335**, 869 (1989).
3, Horn P.S.: J. Amer. Stat. Ass. **78**, 930 (1983).
4. Likeš J., Laga J.: *Základní statistické tabulky*, SNTL Praha 1978.
5. Meloun M., Militký J.: *Statistické zpracování dat*, Plus Praha 1994.
6. Govindaraju K., Rubeška I., Paukert T.: Geost. News. Nancy **18**, 1 (1994).