

## ANNEX 5

# Summarizing urinary iodine data: a worked example

Some actual urinary iodine data from schoolchildren in Cameroon, following the implementation of USI, are presented in the first (left) column of Table 14. The data have been entered into a spreadsheet on a personal computer for ease of calculation. However, with small numbers such as these, the calculations are relatively easily performed by hand.

### A5.1 Steps in processing the data

1. Before proceeding, carefully check the data entered against the original. Ensure that the same number of data points ( $n$ ) is present, and look for any anomalous results.
2. Next, sort the data from highest to lowest, or vice-versa. The spreadsheet will do this automatically.<sup>1</sup> The sorted data are shown under “Value” in Table 14, starting with the highest value, and a summary is shown in Table 15. The next columns show the rank and percentile for each data point.
3. The median is the middle value of the ranked data. In other words, it is the value of the  $(n+1) / 2$ th value. In this case, there are 98 data points, so the median is the value of  $(98+1)$  divided by  $2 = 49.5$ th data point. Accordingly, use the middle point between the 49th and 50th values: 122 and 121  $\mu\text{g/l}$ , respectively. The mid-point is 121.5  $\mu\text{g/l}$ , so the median is 121.5  $\mu\text{g/l}$ .
4. Next, calculate the number of values below 100, 50, and 20  $\mu\text{g/l}$ , respectively. The ranking will allow this to be done very easily. In this case, there are 33 values below 100  $\mu\text{g/l}$ , 6 below 50  $\mu\text{g/l}$ , and one below 20  $\mu\text{g/l}$ . These should be calculated as percentages: 33 of 98 is 33.7%, 5 of 98 is 5.1% and 1 of 98 is 1.0%.
5. Check if any values are above 500  $\mu\text{g/l}$ . There is one (1.0%).
6. The 20th and 80th percentiles may be readily observed, or automatically displayed using the PERCENTILE function [=PERCENTILE

---

<sup>1</sup> In Microsoft Excel, use the Data Analysis function on the Tools menu, and select “Rank and Percentile”.

(range of cells, 0.2)]. The 20th percentile (P20) is 82.4  $\mu\text{g/l}$  and P80 is 191.8  $\mu\text{g/l}$ .

7. The “Descriptive Statistics” function of Data Analysis in Excel provides all statistics shown: select “Summary Statistics” in the dialogue box. Note that the mean is much higher than the median, indicating that the distribution is heavily skewed to the right. This is also shown by the much greater distance between P80 and the median, compared to that between P20 and the median.
8. In addition, the data can be shown as a histogram using the “Histogram” function of Data Analysis in Excel. Convenient ranges need to be chosen for making the frequency distribution, which will be reflected in the height of each bar of the histogram. 50  $\mu\text{g/l}$  is suggested (i.e., the first bar is 0–49  $\mu\text{g/l}$ , the second 50–99  $\mu\text{g/l}$ , the third 100–149  $\mu\text{g/l}$ , etc.). Appropriate modifications can be made using “Chart Options” and related functions. The histogram is shown in Figure 6. A fully detailed description for constructing that histogram is not given here.

In presenting this distribution, it is important not to misinterpret the different percentages. A common mistake is to assume that there is deficiency because 33.7% have a UI value < 100  $\mu\text{g/l}$ . This is not the correct interpretation of the median value and distribution statistics. Instead, this calculation shows the distribution of values around the median value, and helps determine if there is a large proportion with either very low or very high values.

These results indicate that there is no iodine deficiency, and that salt iodization is therefore having the required impact. There is no evidence of significant over-iodization. No changes are needed on the basis of these results, but further follow-up is always essential.

**Table 14 Urinary iodine data in Cameroon schoolchildren following salt iodization**

UI( $\mu\text{g/l}$ )	VALUE	RANK	PERCENT	DESCRIPTIVE	STATISTICS
141	535	1	100.00%		
138	480	2	98.90%	Mean	142.7449
138	395	3	97.90%	Standard error	8.877338
154	340	4	96.90%	Median	121.5
162	320	5	95.80%	Mode	138
26	295	6	94.80%	Standard deviation	87.88117
63	273	7	92.70%	Sample variance	7723.099
111	273	7	92.70%	Kurtosis	5.463542
120	264	9	91.70%	Skewness	1.970291
65	261	10	90.70%	Range	525
190	240	11	89.60%	Minimum	10
142	232	12	87.60%	Maximum	535
138	232	12	87.60%	Sum	13989
95	224	14	86.50%	Count	98
273	208	15	85.50%	Confidence level (95.0%)	17.61905
132	200	16	83.50%		
164	200	16	83.50%		
66	198	18	82.40%		
158	193	19	80.40%		
114	193	19	80.40%		
118	190	21	79.30%		
232	188	22	78.30%		
145	180	23	77.30%		
94	174	24	76.20%		
90	164	25	75.20%		
122	162	26	74.20%		
114	160	27	73.10%		
340	158	28	72.10%		
193	154	29	71.10%		
135	150	30	70.10%		
261	146	31	68.00%		
75	146	31	68.00%		
63	145	33	67.00%		
264	144	34	65.90%		
142	142	35	63.90%		
174	142	35	63.90%		
121	141	37	62.80%		
395	140	38	60.80%		
320	140	38	60.80%		
240	138	40	57.70%		
140	138	40	57.70%		
66	138	40	57.70%		
146	135	43	56.70%		
115	133	44	55.60%		
82	132	45	54.60%		
82	128	46	53.60%		
535	124	47	52.50%		
74	122	48	50.50%		
35	122	48	50.50%	The median lies halfway between	
83	121	50	49.40%	these two values	

**Table 14 continued**

UI( $\mu\text{g/l}$ )	VALUE	RANK	PERCENT	DESCRIPTIVE	STATISTICS
104	120	51	46.30%		
64	120	51	46.30%		
208	120	51	46.30%		
49	118	54	45.30%		
89	117	55	44.30%		
109	115	56	42.20%		
106	115	56	42.20%		
32	114	58	40.20%		
128	114	58	40.20%		
232	111	60	39.10%		
88	110	61	38.10%		
115	109	62	37.10%		
144	108	63	36.00%		
86	106	64	35.00%		
150	104	65	34.00%		
224	96	66	32.90%	<100 $\mu\text{g/l}$	
92	95	67	30.90%		
180	95	67	30.90%		
193	94	69	29.80%		
133	92	70	28.80%		
80	90	71	26.80%		
87	90	71	26.80%		
96	89	73	25.70%		
120	88	74	24.70%		
146	87	75	22.60%		
160	87	75	22.60%		
124	86	77	21.60%		
90	83	78	20.60%		
10	82	79	18.50%		
55	82	79	18.50%		
108	80	81	16.40%		
480	80	81	16.40%		
80	75	83	15.40%		
122	74	84	14.40%		
198	66	85	12.30%		
200	66	85	12.30%		
87	65	87	11.30%		
200	64	88	10.30%		
188	63	89	8.20%		
54	63	89	8.20%		
273	55	91	7.20%		
120	54	92	6.10%		
140	49	93	5.10%	<50 $\mu\text{g/l}$	
110	42	94	4.10%		
42	35	95	3.00%		
95	32	96	2.00%		
117	26	97	1.00%		
295	10	98	.00%	<20 $\mu\text{g/l}$	

**Table 15 Summary of results**

Number	98
Median	121.5 µg/l
20th percentile	82.4 µg/l
80th percentile	191.8 µg/l
<b>Distribution</b>	
<100 µg/l	33.7%
<50 µg/l	5.1%
<20 µg/l	1.0%
>500 µg/l	1.0%

**Figure 6 Frequency table and histogram to show distribution of urinary iodine values after iodization in Cameroon**

URINARY IODINE (µg/l)	FREQUENCY
0-49	6
50-99	27
100-149	35
150-199	13
200-249	7
250-299	5
300-349	2
350-399	1
400-449	0
450-499	1
500-549	1
550-599	0

