

Evaluation of interlaboratory comparison tests & Proficiency Testing (PT) schemes

Leopoldo Cortez & Kees J.M. Kramer

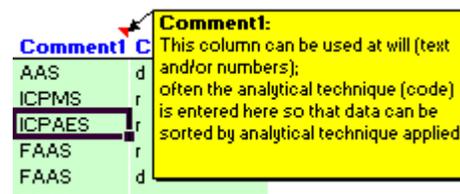
In brief

Tool4PT Pro is a professional tool for the evaluation of interlaboratory comparisons, operating in MS-Excel. Easy copy-paste entry of participant replicates. Handles up to **200 laboratories**, for up to **10 replicates** per participant. Features full statistical evaluation and reporting of results in tables as well as charts, which can be tuned to user demands. Assistance in the use of **Tool4PT Pro** is provided by 'Comments' (hover over a cell), and the worksheet 'Guidance'.

After initial screening of the datasets the definitions of reference value (X_{Ref}) and deviation unit (S_{Ref}) may be based on the 'Classical approach' (outlier testing and exclusion of results) and by 'Robust approach' (by automated Winsorisation). Evaluation of laboratory results is then provided as Z-Scores and En-values. Formats are given for preparation of automatically generated reports to clients.

Compatible with ISO 5725, parts 1-6 (Statistical Methods for Quality Control), ISO 13528 (Statistical methods for use in proficiency testing by interlaboratory comparisons) and ISO Guide 43-1 (Proficiency testing by interlaboratory comparisons. Part 1: Development and operation of proficiency testing schemes). **Tool4PT Pro** has been externally validated.

Operates under Windows 2000 and higher, including Windows XP, Vista and Windows7 64; it requires MS Excel 2000 or higher.



Introduction / Guidance / Report2 / Report1 / Data / Classical / Robust / Evaluation / r & R / B C R /

Outline

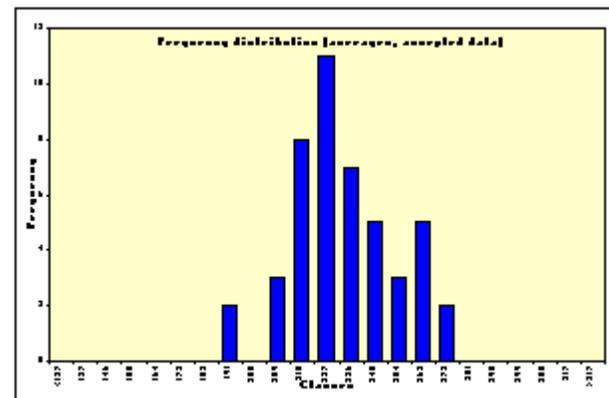
Tool4PT Pro has been designed in 10 worksheets in MS-Excel. In order to minimise data handling and avoiding (typing) errors, all worksheets are fully linked. The worksheet 'Introduction' provides more detail on the tool, and provides instructions on Excel settings (e.g. to enable macros). **Tool4PT Pro** has been developed to be used by professionals who know the business of PT evaluation (e.g. Thomson et al. 2006¹; ISO 17043); references for further study are provided in the product. However, in order to support (new) users the worksheet 'Guidance' provides instructions on the approach best followed when using **Tool4PT Pro**. The other worksheets are discussed below.

Data

This is the first worksheet to be used when a PT evaluation starts: data entry and first screening of the data set for consistency. In order to protect the integrity of the tool, only a limited number of (green) cells accept user input. Data can be entered by typing or copy-paste (preferred: 'Paste Values'); up to 200 PT participants can be entered, each with up to 10 replicates. Additional information may be added: Label (laboratory-code) and laboratory uncertainty (may be left open); Comment1 (e.g. technique) and Comment2 (e.g. extraction or destruction method applied) may help in a technical discussion. For each laboratory n, average, standard deviation, RSD, max, min and median are calculated.

n	Average	St.Dev.	r.s.d.	Max	Min	Median
3	130	35	26.9%	155	90	145
3	162.3333	16.1658	10.0%	177	145	165
3	189	3.60555	1.9%	193	186	188

Accepted/expected data ranges can be set in order to highlight faulty (red) entries (e.g. typing errors). Data can be sorted by Label, Average, Comment1 or Comment2. Charts are automatically generated that support screening of the data set by the user (PT organiser): max-median-min, data replicates and frequency distribution of laboratory means (of raw data & accepted data sets).



¹ Thompson, M., S.L.R. Ellison & R. Wood, 2006. The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories (IUPAC Technical Report) Pure Appl. Chem. 78: 145-196.

Classical approach

The 'Classical approach' worksheet provides the statistical tools to evaluate laboratory results for being a (statistical) outlier. The Cochran test is used for testing of outlying variances. For testing for outlying means the Grubbs single and Grubbs double tests, w/s test and Hampel test are included. The Kolmogorov-Smirnov test is used for testing for normality. For each test outliers and stragglers are flagged. As a result of the statistical evaluation (and PT provider interpretation and its policy) laboratory data sets may be excluded from calculations by entering an 'x' in the reject column. What remains are the 'Accepted data', from which the reference value X_{Ref} and deviation unit s_{Ref} are calculated.

Raw Data					Grubbs outlier		w/s	Hampel	Cochran	Reject
Label	Comment1	n	Average	St.Dev.	Single	Double	test	test	test	
Lab 077	AAS	4	115	41.4327	Straggler		Outlier←	Outlier-		x
Lab 090	ICPMS	10	189.4	59.7033					Outlier	
Lab 084	FAAS	10	206.79	3.99123						
Lab 085	FAAS	10	207.756	6.98903						

Statistics for all data are identified in the 'Raw data' table, the remaining data in the 'Accepted data' table. Full statistical references are provided.

Standard charts show the laboratory mean data with their confidence intervals (to be user defined 80-100%). One may choose to include only the accepted data sets, or to include the rejected means as red dots.

Grubbs Single Outlier Test		
G Critical (1%)	3.060	Outlier
G Critical (5%)	2.758	Straggler
G(Maximum) =	2.391	OK
G(Minimum) =	2.052	OK

Kolmogorov-Smirnov Test		
D Critical (1%)	0.356	
D Critical (5%)	0.284	
D Observed =	0.091	Normal

Grubbs Double Outlier Test		
G Critical (1%)	0.393	Outlier
G Critical (5%)	0.471	Straggler
G(two Max) =	0.608	OK
G(two Min) =	0.705	OK

Statistics of Data Sets			
	Raw data	Accepted	
Nr Data Sets	27	22	
Rejected Sets	5	19%	
Nr Replicates	3	3	
	Raw data	Accepted	
Mean of Means	244.836	247.769	
StDev	40.7217	14.738	
Maximum	350.667	283	
Minimum	130	217.527	
Median	247.667	248.383	
MAD	12.3333	9.23333	

w/s Outlier Test		
w/s Critical (1%)	4.430	Outlier
w/s Critical (5%)	4.790	Straggler
w/s(Max-Min) =	4.442	OK

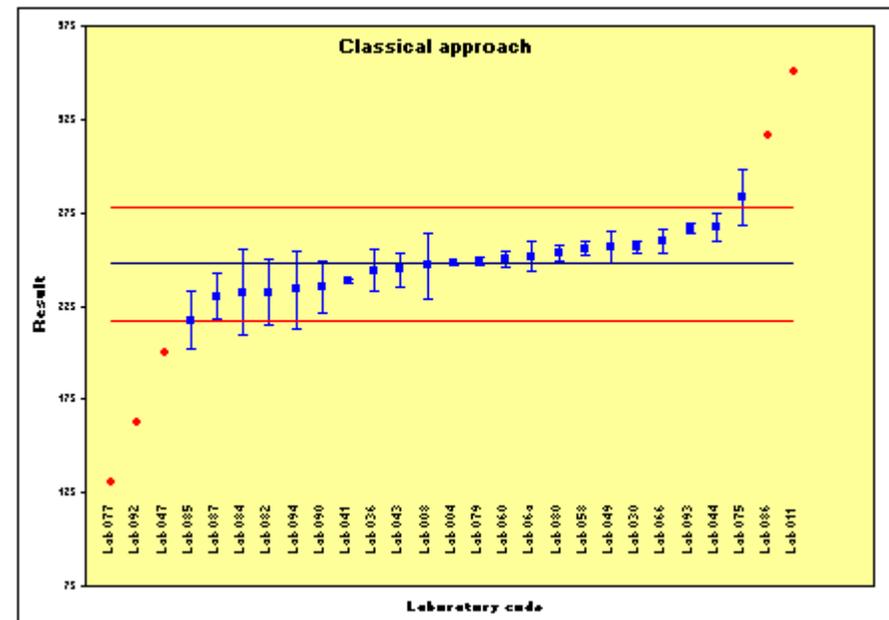
Cochran Outlier Test		
C Critical (1%)	0.307	Outlier
C Critical (5%)	0.252	Straggler
C(Max Var) =	0.184	OK

Statistics for Variance data			
	Raw data	Accepted	
Mean	202.931	131.416	
St.Dev.	266.396	151.363	
Maximum	1225	531.67	
Minimum	0.33333	0.33333	

Hampel Outlier Test		
Hampel Min	206.833	Outlier-
Hampel Max	289.933	Outlier+

Display Rejected in Graph?	Yes
Confidence Interval?	95.00%

All charts in **Tool4PT Pro** can be adjusted using normal MS-Excel operations. Charts can be copy-pasted into other applications, and/or to the 'Report1' worksheet.



Robust approach

The worksheet 'Robust approach' follows the Winsorisation algorithm as mentioned in Analytical Methods Committee², and the Algorithm A in ISO 5725 part 5 to define X_{Ref} and s_{Ref} . A detailed guidance accompanies this function.

Cut-off value (c) and the robust variance constant (β) are set by default at 1.5 and 0.778 respectively, but they may be user defined. After setting the Digit Rounding & Iteration Limit, Winsorisation can be initiated. This may be either by manual step-by-step operation or by an auto-iterate operation.

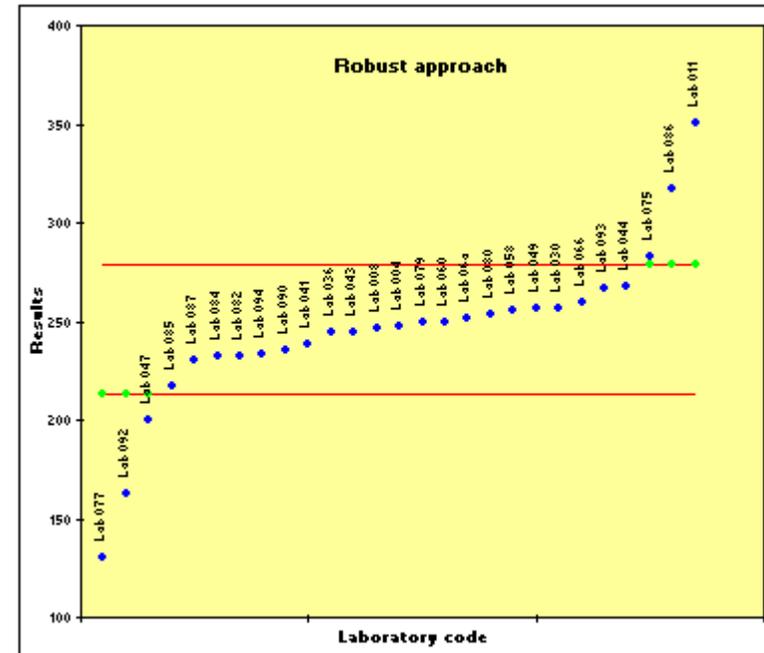
Raw Data		Count =	27
Average =	244.84	Median =	247.67
St.Dev. =	40.73	s(MAD) =	18.29
		MAD =	12.33
Minimum =	130		
Maximum =	350.67		

Winsorised Data		Nr of Iterations:
Nr of Reallocated Data (22%)	6	for Average: 4
		for St.Dev.: 12
Robust Average (H15) =	246.1	
Robust St.Dev. (H16) =	22	
r.s.d. =	9%	
Digit Rounding & Iteration Limit =	0.01	

Winsorisation Parameters			
Cut-Off, c =	1.5	β =	0.778
c <small>(small n)</small> =	1.47	Proposed β =	0.778
Iteration Progress			
Initial Values	Next It	Last It	Difference
246.544259	246.0338	246.03361	2.29E-08
13.4335938	21.38762	21.9370847	0.003464

place the value of reallocation is provided in a table.

The statistics for the 'Raw data' are tabulated together with the 'Winsorised data'. The robust X_{Ref} and s_{Ref} result from the iteration process. The number of iterations and the number and percentage of reallocated data are presented in a table.



In manual operation, the iteration process can be followed as the iteration develops, both in the results table and graphically. The laboratory means are sorted by mean results order. As the Winsorisation algorithm is applied, reallocation of (extreme) data may occur, and the results of each calculation step can be followed in the Winsorisation parameters box. In automatic operation the iteration continues until the iteration limit has been reached.

In the 'Robust approach chart', original data are depicted in blue, reallocated/Winsorised data in green, together with cut-off lines in red. For each laboratory data set where reallocation took

² Analytical Methods Committee, 1989. Robust Statistics - How not to reject outliers. Part 1: Robust Statistics. Analyst, 114: 1693, and: -- Part 2: Inter-laboratory trials. Analyst, 114: 1699.

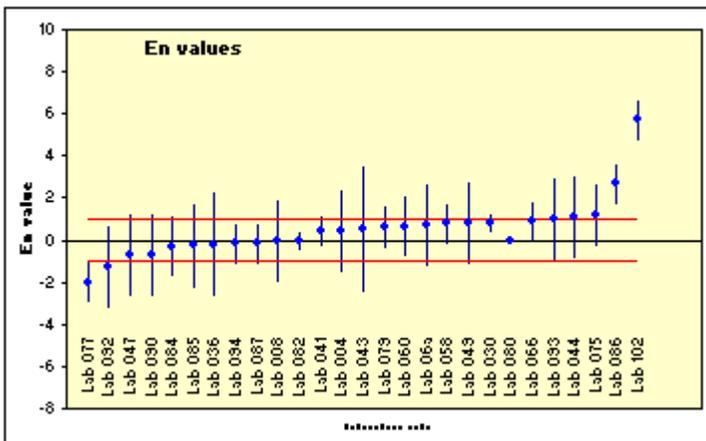
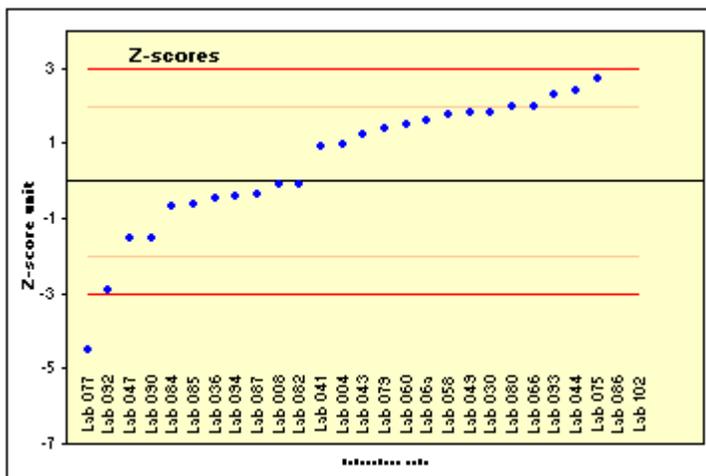
Evaluation

The worksheet 'Evaluation' allows three options for selecting the evaluation criteria X_{Ref} and s_{Ref} : those calculated by the Classical approach, by the Robust approach, or assigned by the user. The latter option allows you to set X_{Ref} and s_{Ref} as a Fit-for-Purpose option (ISO 13528). For the En evaluation the Uncertainty of the Reference Value $U(X_{Ref})$ shall be entered.

The Z-Scores and En-values are computed automatically, using the formulae:

$$Z\text{-Score} = (X_{Lab} - X_{Ref}) / s_{Ref} \quad \text{and} \quad En = (X_{Lab} - X_{Ref}) / \text{SQRT}(U_{Lab}^2 + U_{Ref}^2).$$

Z-Score evaluation identifies $|Z\text{-Score}| > 3$ as unsatisfactory result (pink), and $2 < |Z\text{-Score}| < 3$ as questionable results (yellow).



Reference Parameters

Import from **Classical**

Classical
Robust
User

Reference Parameters

Import from **Classical** bet

User Assigned

XRef: 233.08 200

sRef: 29.47 20

U(XRef): 40

Round digits to nearest:

for Ref.Parameters 0.01

for Lab. Data 0.001

Summary of Statistical conclusions

	Classical	Robust
XRef =	233.08	236.15
sRef =	29.47	36.25
Nr Outliers	3	4
	11.1%	14.8%

The 'Z-Scores' graph shows lines for X_{Ref} , $X_{Ref} \pm 2*s_{Ref}$ and $X_{Ref} \pm 3*s_{Ref}$.

The En evaluation identifies $|En| > 1$ as unsatisfactory results (yellow). The 'En values' graph shows the $X_{Ref} \pm U(X_{Ref})$ lines in red.

All calculated results are presented in a table for each participant.

Label	Comment1	n	Average	St.Dev.	Uncert.	Z-Score	En
Lab 077	AAS	3	130	35	1	-3.5	-2.6
Lab 092	ICPMS	3	162.334	16.166	2	-2.4	-1.8
Lab 047	ICPAES	3	189	3.606	2	-1.5	-1.1
Lab 090	ICPMS	10	189.4	59.704	2	-1.5	-1.1
Lab 084	FAAS	10	206.79	3.992	1.5	-0.9	-0.7
Lab 085	FAAS	10	207.756	6.99	2	-0.9	-0.6

r&R interlaboratory tests

In the worksheet 'r&R interlaboratory test' the Mandel h & k consistency is checked. For each laboratory the 1% and 5% consistency is calculated (ISO 5725 part 2). The 'h' checks on the average value, and 'k' check on the spread/variance value, using 1% and 5% confidence intervals. Suspicious results are flagged and displayed in the table.

Confirmed outliers or stragglers may be rejected by flagging with an 'x' mark on 'Reject' column (the selection may be different from the classical outlier tests).

The 'ANOVA table' displays the interlaboratory variance $s(L)$, the repeatability $s(r)$ and reproducibility $s(R)$ statistics for the accepted dataset.

h&k Charts display the h-consistency and k-consistency, each with lines at 1% and 5%.

Raw Data				Mandel Consistency			
Label	Comment1	Average	St.Dev.	n	h	k	Checks
Lab 077	AAS	130	35	3	-1.817	1.779	k 5%
Lab 092	ICPMS	162.333	16.1658	3	-1.293	0.822	
Lab 047	ICPAES	189	3.60555	3	-0.862	0.183	
Lab 090	ICPMS	189.4	59.7033	10	-0.855	3.034	k 1%
Lab 084	FAAS	206.79	3.99123	10	-0.574	0.203	

[A] Mandel Statistics

h Critical (1%)	2.42	Outlier
h Critical (5%)	1.9	Straggler
k Critical (1%)	2.08	Outlier
k Critical (5%)	1.6	Straggler

[C] ANOVA results

	St.Dev.	Variance
Inter-Lab:	78.4428	6153.27
Repeatability:	17.5026	306.342
Reproducibility:	80.3717	6459.62

[B] Statistics for Average data

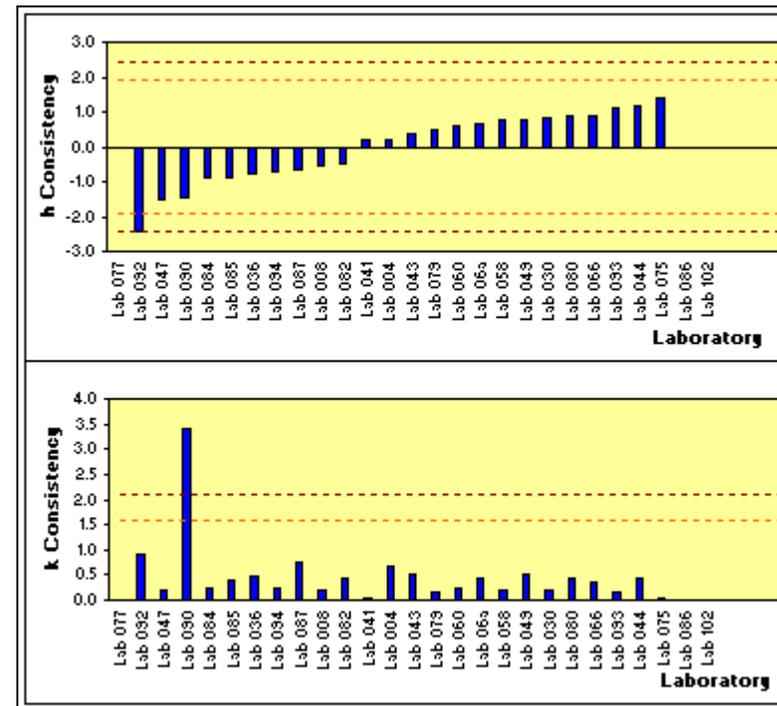
	Raw data	Accepted
Mean Xm	242.247	233.0737
St.Dev.	61.7873	29.46823
Maximum	477.09	274
Minimum	130	162.3333
Count	27	24

[D] Statistics of accepted data

Nr Laboratories, p	24
Nr Replicates, n	3
T1	29435.2
T2	6753793
T3	147
T4	1071
T5	37680.1

Statistics for Variance data

	Raw data	Accepted
Mean S ²	288.599	199.8503
Maximum	3564.49	3564.489
Minimum	0.52	0.52



BCR charts

During the EC-BCR³ programme, evaluation and discussions were conducted using the well known 'Tables and Bar-Charts'. They were very popular as they served well in the technical discussions on accepted data sets, rejection of outliers and on the definition of mean of means and mean of all data. The idea has been copied to the worksheet 'BCR'.

Confidence intervals can be set at 1, 2 or 3s, or at user defined values (e.g. 90%). The confidence intervals are calculated for the accepted data set and reported for the 'Mean of means' and the 'Mean of all data'. The 'BCR'-chart can be set to display the one or the other. Statistics for 'Participant data' (raw data, accepted data), for 'All data', for 'Average data' and for the 'Standard deviation' are provided in a table.

Participants Statistics		
	Raw data	Accepted
Nr DataSets	27	24
Nr Replicates	3	3
Nr DataPoints	147	130

Confidence Interval		
Choose:	2s (95%)	
s		
2s (95%)	5%	
3s (99%)	1%	
	90.0%	10.0%

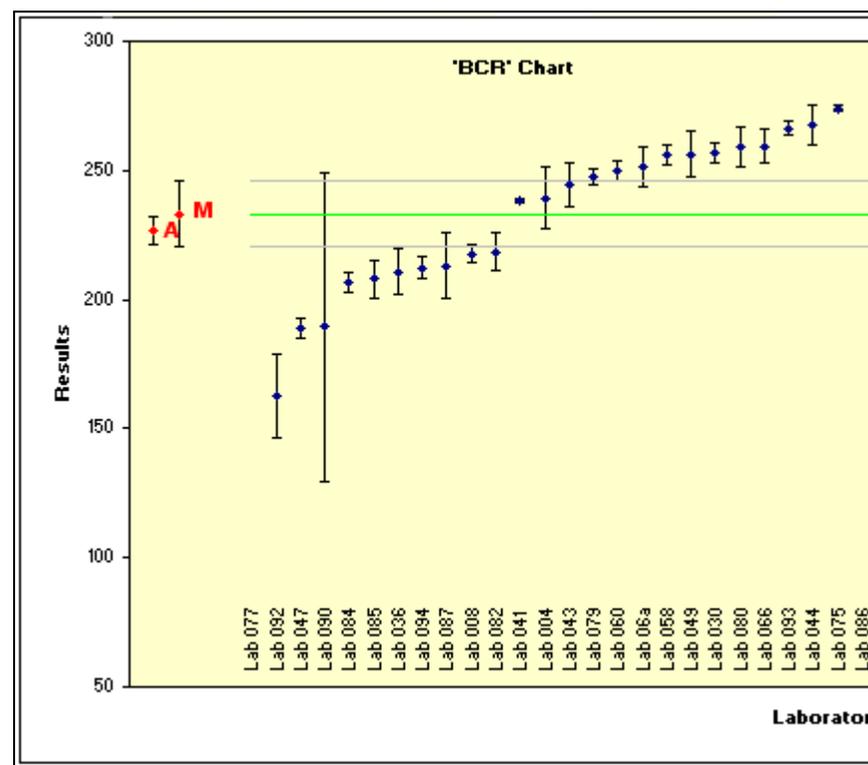
Statistics of All data		
	Raw data	Accepted
Mean All data	244.594	226.4247
St.Dev.	73.7689	31.32902
Count	147	130
Maximum	497.7	275
Minimum	20	20

Accepted data Confidence Interval		
Mean of all data	226.5 ± 5.5	
	C.V. 2.4%	
Mean of means	233.2 ± 12.5	
	C.V. 5.4%	
Significant digits?	4	0.1

Statistics of Average data		
	Raw data	Accepted
Mean Means	242.274	233.1042
St.Dev.	61.7857	29.47092
Maximum	477.1	274
Minimum	130	162.4

Plot Conf.Int. for **Mean of Means**

Statistics for St.Dev. data		
	Raw data	Accepted
Mean S.D.	10.8704	8.554167
Maximum	59.8	59.8
Minimum	0.8	0.8



³ BCR: Bureau Communautaire de Référence, EC Bureau of Reference Materials

Report 1 and 2

Analyte: Analyte1				
Matrix: Matrix1				
Unit: mg/kg				
Statistics of Data				
Classical approach			Robust approach	
	<i>Raw data</i>		<i>Accepted</i>	
Nr Data Sets	27	23	Nr Data Sets	27
Rejected Data		4	Reallocated Data	4
% rejected		14.8%	% reallocated	14.8%
Mean	242.247	234.98	Robust Mean	236.15
St.Dev.	61.788	28.59	Robust St.Dev	36.25
C.V.(%)	25.5	12.2	C.V.(%)	15.4
Maximum	477.09	274	Iteration limit used	0.001
Minimum	130	162.334	Cut-off value, c used	1.5
Median	244.334	244.334	Beta used	0.778
Z-score Evaluation based on:				
Definition of Xm and S:	User defined			
Reference value	XRef	220		
Deviation unit	sRef	20		

The worksheet 'Report 1' contains a template that will facilitate the reporting of the results. In contrast to the other worksheets this sheet may be tuned to the demands of the user.

The first page contains the synthesis of the outcome of the evaluation: results from the 'Classical approach' and the 'Robust approach'. It is indicated on what criterion the evaluation has been made. The second and following pages provide the participants' results with all key results (n, average, RSD, Z-score and En value). Unsatisfactory and questionable results are flagged. The 'Comment2' is copied from the worksheet 'Data'.

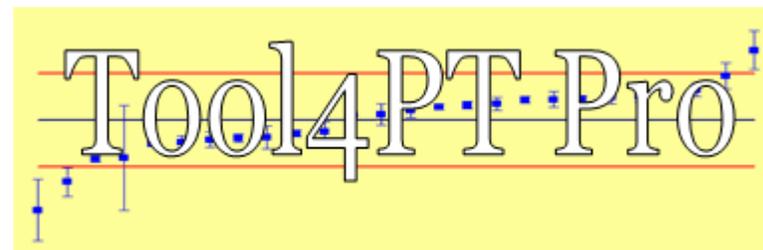
The Report1 provides a template for the charts that need to be reported to the PT participants: Classical (all, accepted) or Robust, Z-Scores and En values (all, accepted).

In the worksheet 'Report2' a full summary of all statistical results is provided in one row for the raw and accepted data, for all relevant tests. This report will be useful when more than one analyte is to be reported: result rows can then be copy-pasted into one overview table.

Analyte: Analyte1										
Matrix: Matrix1										
Unit: Unit1										
Results by Laboratory										
<i>Label</i>	<i>Comment1</i>	<i>n</i>	<i>Average</i>	<i>St.Dev.</i>	<i>RSD</i>	<i>Z-Score</i>	<i>"I"</i>	<i>En</i>	<i>"</i>	<i>Comment2</i>
Lab 077	AAS	3	130.00000	35.00000	26.9%	-4.5	**	-2.0	**	d
Lab 092	ICPMS	3	162.33400	16.16600	10.0%	-2.9	*	-1.3	**	r
Lab 047	ICPAES	3	189.00000	3.60600	1.9%	-1.6		-0.7		r
Lab 090	ICPMS	10	189.40000	59.70400	31.5%	-1.5		-0.7		r
Lab 084	FAAS	10	206.79000	3.99200	1.9%	-0.7		-0.3		f

Performance

Operating **Tool4PT Pro** is easy. It operates in MS-Excel, a spreadsheet program that most scientists use every day. In order to avoid their disruption, worksheets are protected (except those for reporting). Nevertheless, functions in Excel remain working as usual. For example, the user can adapt the graph's x and y axis scales or copy-paste tables. No need to learn this.



The tool itself comes with assistance where you need it (hover over cells to see the 'Comment'). There is a Guidance included, which outlines the sequence of actions. The use of **Tool4PT Pro** starts with entering/copy-pasting data, it performs the full statistics and assists you to evaluate the results, and it ends with the reporting sheets template already filled in. Still it leaves sufficient room for the user to set individual interpretation. A full evaluation run takes 15-20 min, reporting has become easy.

Validation

ASLAB⁴ tested **Tool4PT Pro** (version 1.04.01) against 10 PT data set evaluations. Dr I. Koruna concluded: "Tool4PT is a well worked-up programme for evaluating PT tests under standardized procedures and its results are fully compatible even under extreme conditions, which appeared in ASLAB PT runs used as testing data sets".

Tool4PT Pro was used in the CoEPT project for the Water and Food sector by the Portuguese PT provider. From the report summaries: "the agreement between the protocols of 10 PT providers in the water sector, tested on drinking water datasets is good", and "The agreement between the results from the five PT providers in the food sector, tested on milk powder datasets, is good"⁵

Requirements

Tool4PT Pro consists of an Excel file supplied on a CD-ROM. Windows 2000 or higher, including Windows XP, Vista 32, Windows7 64; Microsoft Excel 2000 or higher is required. After evaluation, files are typical 2-2.5 MB in size.

Further information

Tool4PT Pro is produced and distributed by: MERMAYDE

Zakedijkje 60, NL – 1862 HC Bergen, Netherlands

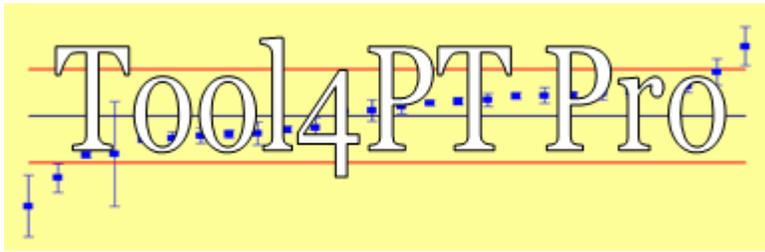
info@mermayde.nl

www.mermayde.nl



⁴ ASLAB - Centre for assessment of laboratories - T. G. Masaryk Water Research Institute, Praha, Czech Republic

⁵ Comparability of the operation and evaluation protocols of European proficiency testing schemes (CoEPT): <http://www.eptis.bam.de/index.html?coept.htm>



Order Form

Please send me _____ copy(ies) of the CD-ROM:

Leopoldo Cortez & Kees J.M. Kramer

Tool4PT Pro: Evaluation of interlaboratory comparisons

MERMAYDE, Bergen, Netherlands, © 2005-2012.

Price € 495.00 (excluding 21% VAT *) and shipping)

Payment is required on receipt, in Euro (€) by transfer to our bank (for payment by cheque: add € 12.50). For credit card payments see below. An invoice will be sent with the shipment.

Name:

Company/Institution:

Address:

Postal Code:

City and Country:

E-mail:

Your purchase order / reference:

(For EC customers) Your VAT Nr *):

Date:

Signature:

*) Value added tax: we have to charge VAT at 21% if you are located in the Netherlands, or if you are in the European Community but are not registered for VAT. VAT is not chargeable on sales outside the European Community. If you are in the European Community, and registered for VAT, then please quote your VAT registration number when ordering.

We accept credit card payments via PayPal:



Please print and send by fax or postal service to:
MERMAYDE, Zakedijkje 60, NL - 1862 HC Bergen - Fax: +31-72.581 56 15