COMPARATIVE STUDY ON THE PROCEDURES FOR REJECTION OF OUTLIERS IN ASCOPE LABORATORY TEST CORRELATION PROGRAMME*

By : E. Jasjfi

ABSTRACT

Grubbs T-Test has been consistently used in statistical analysis for rejection of outliers in ASCOPE Laboratory Test Correlation Programme.

The present paper compares the effects of using other statistical tests, namely Ferguson b_2-Test and Dixon r-Test on the rejection of laboratory test results.

Although these methods gave conflicting verdicts on some of the laboratory results, it seems that the use of Grubbs T-Test is quite satisfactory for the purpose of ASCOPE Laboratory Test Correlation Programme.

I. INTRODUCTION

ASCOPE Laboratory Test Correlation Programme is a cooperative programme among petroleum testing laboratories in ASEAN countries. The programme has been going on since 1979, and currently comprises three sets of programmes namely for (1) lubricating oils, (2) fuels, and (3) CFR engines [5]. Each set is carried out twice a year so that presently, for the year 1990, the programmes are the 18th and 19th for lubricating oils, the 20th and 21st for fuels, and the 17th and 18th for CFR engines.

Two samples are generally tested each time, namely fresh oil and used oil for the lube correlations, jet fuel and diesel fuel for the fuel correlation, and regular and premium grade gasoline for the CFR engine correlation programme.

Since the first of these programmes, ASCOPE laboratory group has been employing Grubbs T-test for detecting and rejection of outliers. ASTM Standard Practice for Dealing with Outlying Observation (ASTM E-178) recognizes several methods for this purpose, of which Grubbs T-test as practised by ASCOPE is one of the recommended methods. Such method is also used by laboratory correlation programmes such as those of Ethyl Corp. and ASTM National Exchange Group which employed it for their CFR engines correlation programmes. Other laboratory groups conducting similar correlation programmes are known to use various other methods. For example, Mobil Oil is using the b_2-test of Ferguson for its correlation programme whereas UNITAR/UNDP uses Dixon *r*-test criteria for detecting outliers in its cooperative programme for developing standard test methods for evaluating heavy crudes.

* *This paper was presented at the 11th Annual ASCOPE Laboratory Workshop, Bandar Seri Begawan, Brunei Darussalam, August 20–22 1990.*

It is therefore of interest to compare the methods and examine the effects of the various methods if applied to ASCOPE laboratory test correlation data It is expected from this comparative study to establish whether to reaffirm the use of the method presently practised by ASCOPE or to make new recommendation concerning the most suitable method for future ASCOPE correlation programme if necessary.

II. METHODS FOR DETECTING OUTLIERS

The ASTM Standard Practice for Dealing with Outlying Observation (ASTM E-178) deals with the problem of outlying observation in samples and the procedures to test the statistical significance of them. An "outlier" is defined as the observation data which appears to deviate markedly from other members of the sample. Such outlier may be just an extreme manifestation of the random variability of the data and does belong to the same population.

On the other hand, it may has arisen from certain deviation in experimental procedure or from error in recording and calculating numerical value. Such data does not belong to the same population as other members of the sample and should be rejected. In fact, if a deviation or an error is known to have been committed in obtaining a data, the data must be rejected no matter how good it agrees with the rest of the data in the sample.

However, if no error is known to have been made in obtaining the data, but it is markedly different from the rest in the sample, a statistical test should be conducted to establish whether such data belong to the same population and should be retained, or must be rejected since it must have came from different population. have came from different population.

ASTM Standard Practice E-178 provides several statistical methods for testing the suspected outliers, of which three will be described here as they will be applied to our ASCOPE data to examine the variability of their results. These are T-test b_2-test, and Dixon r-test.

Basically, in all these methods, the suspected data is included in the initial calculation of the numerical value of a sample criterion (or statistic). This value is then compared with a critical value as established in each of the methods to determine whether the suspected value should be rejected or retained. The critical value is the value that would be exceeded by chance with some specified small probability on the assumption that all the observations did indeed constitute a random sample of a single parent population.

The specific probability is called the "significance level" and indicates the risk of erroneously rejecting a good observation (or making Type I error). If there exists a real shift in the value of the observation that arises from non-random causes (human error, loss of calibration, change of measuring instrument, or even time of measurement, etc.) then the observed value of the sample criterion used would exceed the critical value based on the random sample theory.

The significance level is usually taken as 1% or 5%, meaning there is a risk of 1 in 100 or 1 in 20, respectively, of erroneously rejecting good observation. Another way of expressing the risk is in "confidence limit", meaning probability of making correct judgement in rejecting the an observation. 1% or 5% significance level is equivalent with 99% or 95%, respectively, confidence limit.

A. Grubbs T-Test

The T-test comparises a test using the critical value of T as proposed initially by Grubbs in 1950 and further extended by Grubbs and Beck in 1972 [4]. The extended version is quoted by, and used in, ASTM E-178.

For this test let the sample of n observations be denoted in order of increasing magnitude by $x_1 \leq x_2 \leq x_3 \leq \ldots < x_n$. Suppose the largest value The test criterion recommended by Grubbs is :

$$T_n = (x_n - \bar{x}) / s$$

where $\bar{x}$ is the arithmetic average of all n values, and s is the standard deviation based on the sample data. Thus

$$\bar{x} = (\Sigma x_1) / n$$
$$s = (\Sigma (x_1 - \bar{x})^2 / (n-1)]^{1/2}$$

If the smallest value, x_1, is the doubtful one, then the Grubbs criterion is

$$T_1 = (\bar{x} - x_1) / s$$

B. Ferguson b_2–Test

Grubbs method may have problems if several outliers are present in the sample since the detection of one or two spurious values may be "masked" by the presence of other anomalous observations. Outlying observation occurs due to a shift in level (or mean), or change in scale (i.e., change in variance of the observations), or both. Ferguson [2] has studied the various rejection rules relative to changes in level or scale. For several outliers and repeated rejection of observations, he suggested the use of the sample coefficient of kurtosis for "two-sided" test change in level to higher and lower values).

The sample coefficient of kurtosis is

$$b_2 = n\Sigma (x_i - \bar{x})^4 / (n-1)^2 s^4$$
$$= n\Sigma (x_i - \bar{x})^4 / [\Sigma (x_i - x)^2]^2$$

C. Dixon r-Test

The Dixon criteria is based entirely on the ratios of distance between a suspected observation and its nearest or next-nearest (assumed unaffected) neighbour to the range of the sample. This is useful where quick judgement is required and it is desirable to avoid calculation of the standard deviation. Dixon suggested the use of the following criteria[1]

For $3 \leq n \leq 7$:

$r_{10} = (x_2 - x_1) / (x_n - x_1)$ if smallest value is suspected

$= (x_n - x_{n-1}) / (x_n - x_1)$ if largest value is suspected.

For $8 \leq n \leq 10$:

$r_{11} = (x_2 - x_1) / (x_{n-1} - x_1)$ if smallest value is suspected

$= (x_n - x_{n-1}) / (x_n - x_2)$ if largest value is suspected.

For $11 \leq n \leq 13$:

$r_{21} = (x_3 - x_1) / (x_{n-1} - x_1)$ if smallest value is suspected

$= (x_n - x_{n-2}) / (x_n - x_2)$ if largest value is suspected.

For $14 \leq n \leq 30$:

$r_{22} = (x_3 - x_1) / (x_{n-2} - x_1)$ if largest value is suspected.

$= (x_n - x_{n-2}) / (x_n - x_3)$ if largest value is suspected.

III. APPLICATION TO ASCOPE DATA

The three different procedures above are applied in this study to the data of 1989 ASCOPE Laboratory Test Correlation Programmes (ALTCP). These are :

(1) 17th ALTCP for Lube Oils: Fresh Oil sample (ALTCP Data A)
(2) 17th ALTCP for Lube Oils: Used Oil sample (ALTCP Data B)
(3) 18th ALTCP for Fuels: Jet Fuel sample (ALTCP Data C)
(4) 18th ALTCP for Fuels: Diesel Fuel sample (ALTCP Data D)
(5) 19th ALTCP for Fuels: Jet Fuel sample (ALTCP Data E)
(6) 19th ALTCP for Fuels: Diesel Fuel sample (ALTCP Data F).

To ilustrate, details for seven test items of ALTCP Data E (19th ALTCP for Fuels, Jet Fuel sample) is presented in Table 1.

For each ALTCP Data the following statistical tests are applied:

(1) Grubbs T-test at 95% confidence limit
(2) Grubbs T-test at 99% confidence limit
(3) Ferguson b_2-test at 95% confidence limit
(4) Ferguson b_2-test at 99% confidence limit
(5) Dixon r-test at 95% confidence limit
(6) Dixon r-test at 99% confidence limit.

Parts of the results for ALTCP Data E (19th ALTCP for Fuels, Jet Fuel Sample) at 95% confidence limit were presented for T-test (Table 2), b_2-test (Table 3), and r-test (Table 4). For the complete set of ALTCP Data E, T-test and b_2-test resulted in six data sets (test methods) with rejection of lab results (Data Sets E.3, E.4, E.7, E.9, E.12 and E.14), while r-test had five data sets with rejection (Data Sets E.3, E.4, E.5, E.7, and E.12).

Closer examination of the complete results revealed that disagreement in rejection by the three statistical tests were observed on three data sets (Data Sets E.5, E.9 and E.14) at 95% confidence limit and four data sets (Data Sets E.3, E.5, E.7, and E.12) for 99% confidence limit. The conflicting results for Data Sets E3 and E5 were illustrated in Tables 5 and 6.

The conflicting verdicts were suffered by Lab-22 and Lab-18 in Data Set E.5 at 95% confidence limit, and Lab-4 and Lab-22 in Data Sets E.3 and E.5, respectively, at 99% confidence limit. In Data Set E.5, for example, the two lab-results were rejected by r-test but not by T-test nor by b_2-test. For Data Set E.9, on the other hand, two test results were rejected by both T-test and b_2-test, but not by r-test.

Such processing and examination were carried out for all ALTCP Data (A through F) and the results were summarized in Table 7 for 95% confidence limit. Table 8 shows the results for 95% and 99% confidence limits, expressed in percentage.

IV. DISCUSSION AND CONCLUSION

Examination of the results shows that, while there are differences in results of the three statistical treatments, the differences were small. Out of a total of 116 test methods involved, application of T-test resulted in 46 test methods with rejection of results, while b_2-test has 47 and r-test also has 47, at 95% confidence level. In number of lab results rejected, they were also very close, namely, 59, 60 and 64 lab results were rejected by T-test, b_2-test and r-test, respectively.

The number of data sets with disagreement in results were small, 11 out of 116 at 95% confidence limit, and the number of lab-results which are subject to conflicting verdict were even smaller, namely 36 lab-results out of 1399 or 2,57% at 95% confidence limit. The figures were even smaller, 1,57%, at 99% confidence limit.

Based on the data above, it can be concluded that although some lab-results may be subject to conflicting verdicts, the number is very small, and no significant difference would be observed in the three statistical tests examined.

It seems that practicality would be the major factor in deciding which statistical test to be selected in evaluating a set of laboratory data. Dixon r-test, for example, would be most practical if one wants to avoid calculation of standard deviation or where quick judgement is called for. Ferguson b_2-test is recommended by ASTM for repeated rejection since the use of coeficient of skewness or coeficient of kurtosis coild avoid the "masking effect" of several outliers which are present at the same time.

For ASCOPE, it seems there is no reason to quit using the Grubbs T-test, which has been employed for so long a time. Our purpose in ASCOPE laboratory test correlation is to assist

participating laboratories to know their performance as compared to others and as compared to average. If one's result is far from the average, it is a good idea to examine one's practice in performing the laboratory test. If one's results tends to be always on the high side or the low side of the average, then the need for self examination becomes more urgent.

The ASCOPE Annual Laboratory Workshop, where members of the participating laboratories meet to discuss the results would be a practical forum where participants could assist each other in finding the reasons of any discrepancy in his laboratory test results.

REFERENCES

1. **Dixon W.J.**, "Processing Data for Outliers", *Biometrics*, Vol. 9, No. 1, March 1953.
2. **Ferguson, T.T.**, "Rules for Rejection of Outliers", *Revue Inst. Int. de Stat.*, Vol. 3, 1961.
3. **Grubbs, F.E.**, "Procedure for Detecting Outlying Observations in Samples", *Technometrics*, Vol. 11, No. 4, February 1969.
4. **Grubbs, F.E.** and **Beck, G.**, "Extension of Sample Sizes and Percentage Points for Significance Test of Outlying Observations", *Technometrics*, Vol. 14, No. 4, November 1972.
5. **Jasjfi, E.**, "ASCOPE Cooperation in Laboratory Test Correlation Programmes", *Proceedings*, Third ASCOPE Conference and Exhibition, Kuala Lumpur, December 2-5, 1985.
6. **Nater, J., Wasserman, W., and Whitmore, G.A.**, "Applied Statistics", 3rd edn., Allyn and Bacon, Inc., Boston, 1988.
7. **Tietjen, G.L. and Moore, R.H.**, "Some Grubbs-Type Statistics for Detection of Outliers", *Technometrics*, Vol. 14, No. 3, August 1972.
8. **Walpole, R.E.** and **Myers, R.H.**, Probability and Statistics for Engineers and Scientists", 2nd., Mac Millan Publishing Co., Inc., New York, 1978.

APPENDIX

Table 1

ALTCP Data E

Data from 19th ALTCP for Jet Fuels - Jet Fuel Sample

Data Set		E.1	E.2	E.3	E.4	E.5	E.6	E.7
No.	Test Lab.	Density a 15°C D-1298	SG at 60/60°F D-1298	Amiline Pt. °F D-611	Flash Pt. Abel °C IP-170	Freezing Point °C D-2386	Smoke Pt.mm IP-57	Aromat % vol. D-1319
1.	A	0.8023	0.8027	63.90	43.0	-52.0	24.5	17.6
2.	B	0.8018	0.8022	61.90	43.0	-50.0	25.5	13.7
3.	C	– –	– –	– –	– –	– –	– –	– –
4.	D	0.8003	0.8007	66.20	39.0	-52.0	25.0	– –
5.	E	0.8023	0.8027	62.50	44.5	– –	21.6	15.1
6.	F	– –	– –	– –	– –	– –	– –	– –
7.	G	0.8016	0.8019	61.60	41.5	-52.0	25.0	17.5
8.	H	0.8018	– –	63.40	43.0	-50.5	24.0	15.3
9.	I	0.8018	0.8017	64.50	44.5	-51.5	24.0	15.8
10.	J	0.8026	0.8030	– –	43.0	– –	– –	– –
11.	K	0.8013	0.8017	62.00	42.0	-53.0	25.0	– –
12.	L	0.8018	0.8022	61.95	40.5	-52.5	25.0	– –
13.	M	0.8021	– –	– –	43.3	-51.0	– –	– –
14.	N	0.8018	– –	62.00	40.0	-52.0	23.0	16.0
15.	O	0.8018	0.8022	62.80	45.0	-51.0	25.0	14.9
16.	P	0.8027	0.8031	63.00	43.0	-48.0	26.0	14.7
17.	Q	– –	– –	– –	– –	– –	– –	– –
18.	R	0.8004	0.8004	63.50	43.3	-47.5	20.0	14.0
19.	S	0.8010	0.8014	62.20	44.0	-52.0	21.0	15.0
20.	T	0.8008	0.8012	62.50	44.5	-50.0	22.0	14.6
21.	U	0.8013	0.8017	165.80	44.0	-53.0	23.0	13.5
22.	V	0.8022	0.8027	91.00	45.0	-56.0	22.0	– –
23.	W	– –	0.8020	157.00	116.0	– –	21.0	9.8
24.	X	– –	– –	– –	– –	– –	– –	– –
25.	Y	0.8014	0.8017	62.60	43.0	-51.0	23.0	– –

Table 2
Results of T-test at 95% Confidence Limit for
ALTCP Data E

Data Set	E.1	E.2	E.3	E.4	E.5	E.6	E.7
Test	Density a 15°C D-1298	SG at 60/60°F D-1298	Aniline Pt. °F D-611	Flash Pt. Abel °C IP-170	Freezing Point °C D-2386	Smoke Pt.mm IP-57	Aromat % vol. D-1319
Number of lab results reported	20	18	19	21	18	19	14
Minimum	0.8003	0.8004	61.6	39	−56.0	20	9.8
Maximum	0.8027	0.8031	165.8	116	−47.5	26	17.6
Number of lab results rejected	0	0	4	1	0	0	1
Lab results rejected	None	None	Lab 21 Lab 23 Lab 22 Lab 4	Lab 23	None	None	Lab 23
After rejection of outliers:							
Minimum	0.8003	0.8004	61.6	39.0	−56.0	20.0	13.5
Maximum	0.8027	0.8031	64.5	45.0	−47.5	26.0	17.6
Average	0.8017	0.8020	62.7	43.0	−51.4	23.5	15.2
Average deviation	0.0005	0.0006	0.7	1.0	1.3	1.5	0.9
Standard deviation	0.0006	0.0007	0.8	2.0	1.8	1.7	1.2
T-factor	2.557	2.504	2.705	2.557	2.504	2.532	2.331
Lower rejection value	0.8000	0.8001	60.8	38.9	−56.0	19.1	12.4
Upper rejection value	0.8033	0.8038	64.6	47.1	−46.8	27.8	18.1

Table 3
Results of b_2 test at 95% Confidence Limit for ALTCP Data E

Data Set	E.1	E.2	E.3	E.4	E.5	E.6	E.7
	Density a 15°C D-1298	SG at 60/60°F D-1298	Aniline Pt. °F D-611	Flash Pt. Abel °C IP-170	Freezing Point °C D-2386	Smoke Pt.mm IP-57	Aromat % vol. D-1319
Number of lab results reported	20	18	19	21	18	19	14
Minimum	0.8003	0.8004	61.6	39	−56.0	20.0	9.8
Maximum	0.8027	0.8031	165.8	116	−47.5	26.0	17.6
Number of lab results rejected	0	0	4	1	0	0	1
Lab results rejected	None	None	Lab 21 Lab 23 Lab 22 Lab 4	Lab 23	None	None	Lab 23
After rejection of outliers:							
Minimum	0.8003	0.8004	61.6	39.0	−56.0	20.0	13.5
Maximum	0.8027	0.8031	64.5	45.0	−56.0	26.0	17.6
Average	0.8017	0.8020	62.7	43.0	−51.4	23.5	15.2
Average deviation	0.0005	0.0006	0.7	1	1.3	1.5	0.9
Standard deviation	0.0006	0.0007	0.8	2	1.8	1.7	1.2
b_2 calculated	2.62	2.58	2.61	3.16	4.01	1.93	2.66
b_2 critical	4.12	4.07	4.15	4.12	4.12	3.98	4.09

Table 4
Results of *r*-test at 95% Confidence Limit for ALTCP Data E

Data Set	E.1	E.2	E.3	E.4	E.5	E.6	E.7
Test	Density a 15°C D-1298	SG at 60/60°F D-1298	Aniline Pt. °F D-611	Flash Pt. Abel °F IP-170	Freezing Point °C D-2386	Smoke Pt.mm IP-57	Aromat % vol. D-1319
Number of lab results reported	20	18	19	21	18	19	14
Minimum	0.8003	0.8004	61.6	39.0	–56.0	20.0	9.8
Maximum	0.8027	0.8031	165.8	116.0	–47.5	26.0	17.6
Number of lab results rejected	0	0	4	1	2	0	1
Lab results rejected	None	None	Lab 21 Lab 23 Lab 22 Lab 4	Lab 23	Lab 22 Lab 18	None	Lab 23
After rejection of outliers:							
Minimum	0.8003	0.8004	61.6	39.0	–53.0	20.0	13.5
Maximum	0.8027	0.8031	64.5	45.0	–48.0	26.0	17.6
Average	0.8017	0.8020	62.7	43.0	–52.3	23.5	15.2
Average deviation	0.0005	0.0006	0.7	1.0	1.0	1.5	0.9
Standard deviation	0.0006	0.0007	0.8	2.0	1.2	1.7	1.2
r-lower	0.250	0.444	0.184	0.273	0.167	0.400	0.125
r-higher	0.211	0.474	0.392	0.111	0.444	0.250	0.410
Dixon criteria	0.450	0.475	0.525	0.450	0.507	0.462	0.521

Table 5

Data Sets with Conflicting Verdicts
by the Three Statistical Procedures

95% Confidence Limit					99% Confidence Limit				
E.3					E.3				
Aniline Pt. °F, D-611					Aniline Pt. °F, D-611				
No.	T test	b_2 test	r test	Lab.	No.	T test	b_2 test	r test	Lab.
1	61.60	61.60	61.60	Lab– 7	1	61.60	61.60	61.60	Lab– 7
2	61.90	61.90	61.90	Lab– 2	2	61.90	61.90	61.90	Lab– 2
3	61.95	61.95	61.95	Lab–12	3	61.95	61.95	61.95	Lab–12
4	62.00	62.00	62.00	Lab–11	4	62.00	62.00	62.00	Lab–11
5	62.00	62.00	62.00	Lab–14	5	62.00	62.00	62.00	Lab–14
6	62.20	62.20	62.00	Lab–19	6	62.20	62.20	62.20	Lab–19
7	62.50	62.50	62.50	Lab– 5	7	62.60	62.50	62.50	Lab– 5
8	62.50	62.50	62.50	Lab–20	8	62.50	62.50	62.50	Lab–20
9	62.60	62.60	62.60	Lab–25	9	62.60	62.60	62.60	Lab–25
10	62.80	62.80	62.80	Lab–15	10	62.80	62.80	62.80	Lab–15
11	63.00	63.00	63.00	Lab–16	11	63.00	63.00	63.00	Lab–16
12	63.40	63.40	63.40	Lab– 8	12	63.40	63.40	63.40	Lab– 8
13	63.50	63.50	63.50	Lab–18	13	63.50	63.50	63.50	Lab–18
14	63.90	63.90	63.90	Lab– 1	14	63.90	63.90	63.90	Lab– 1
15	64.50	64.50	64.50	Lab– 9	15	64.50	64.50	64.50	Lab– 9
16	*66.20*	*66.20*	*66.20*	*Lab– 4*	16	*66.20*	66.20	66.20	*Lab–4**
17	*91.00*	*91.00*	*91.00*	*Lab–27*	17	*91.00*	*91.00*	*91.00*	*Lab–22*
18	*157.00*	*157.00*	*157.00*	*Lab–23*	18	*91.00*	*91.00*	*157.00*	*Lab–23*
19	*165.80*	*165.80*	*165.80*	*Lab–21*	19	*165.80*	*165.00*	*165.00*	*Lab–21*
20	– –	– –	– –	*Lab– 3*	20	– –	– –	– –	*Lab– 3*
21	– –	– –	– –	*Lab– 6*	21	– –	– –	– –	*Lab– 6*
22	– –	– –	– –	*Lab–10*	22	– –	– –	– –	*Lab–10*
23	– –	– –	– –	*Lab–13*	23	– –	– –	– –	*Lab–13*
24	– –	– –	– –	*Lab–17*	24	– –	– –	– –	*Lab–17*
25	– –	– –	– –	*Lab–24*	24	– –	– –	– –	*Lab–24*

*Conflicting verdicts.
Italic means result rejected.

Table 6

Data Sets with Conflicting Verdicts
by the Three Statistical Procedures

95% Confidence Limit					99% Confidence Limit				
E.5					E.5				
Freezing Point °C, D-2386					Freezing Point °C, D-2386				
No.	T test	b_2 test	r test	Lab.	No.	T test	b_2 test	r test	Lab.
1	–56.0	–56.0	*–56.0*	*Lab–22**	1	–56.0	–56.0	*–56.0*	*Lab–22**
2	–53.0	–53.0	–53.0	Lab–11	2	–53.0	–53.0	–53.0	Lab–11
3	–53.0	–53.0	–53.0	Lab–21	3	–53.0	–53.0	–53.0	Lab–21
4	–52.5	–52.5	–52.5	Lab–12	4	–52.5	–52.5	–52.5	Lab–21
5	–52.0	–52.0	–52.0	Lab– 1	5	–52.0	–52.0	–52.0	Lab– 1
6	–52.0	–52.0	–52.0	Lab– 4	6	–52.0	–52.0	–52.0	Lab– 4
7	–52.0	–52.0	–52.0	Lab– 7	7	–52.0	–52.0	–52.0	Lab– 7
8	–52.0	–52.0	–52.0	Lab–14	8	–52.0	–52.0	–52.0	Lab–14
9	–52.0	–52.0	–52.0	Lab–19	9	–52.0	–52.0	–52.0	Lab–19
10	–51.5	–51.5	–51.5	Lab– 9	10	–51.5	–51.5	–51.5	Lab– 9
11	–51.0	–51.0	–51.0	Lab–13	11	–51.0	–51.0	–51.0	Lab–13
12	–51.0	–51.0	–51.0	Lab–15	12	–51.0	–51.0	–51.0	Lab–15
13	–51.0	–51.0	–51.0	Lab–25	13	–51.0	–51.0	–51.0	Lab–25
14	–50.5	–50.5	–50.5	Lab– 8	14	–50.5	–50.5	–50.5	Lab– 8
15	–50.0	–50.0	–50.0	Lab– 2	15	–50.0	–50.0	–50.0	Lab– 2
16	–50.0	–50.0	–50.0	Lab–20	16	–50.0	–50.0	–50.0	Lab–20
17	–48.0	–48.0	–48.0	Lab–16	17	–48.0	–48.0	–48.0	Lab–16
18	–47.5	–47.5	*–47.5*	*Lab–18**	18	–47.5	–47.5	–47.5	Lab–18
19	– –	– –	– –	*Lab– 3*	19	– –	– –	– –	*Lab– 3*
20	– –	– –	– –	*Lab– 5*	20	– –	– –	– –	*Lab– 5*
21	– –	– –	– –	*Lab– 6*	21	– –	– –	– –	*Lab– 6*
22	– –	– –	– –	*Lab–10*	22	– –	– –	– –	*Lab–10*
23	– –	– –	– –	*Lab–17*	23	– –	– –	– –	*Lab–17*
24	– –	– –	– –	*Lab–23*	24	– –	– –	– –	*Lab–23*
25	– –	– –	– –	*Lab–24*	25	– –	– –	– –	*Lab–24*

*Conflicting verdicts

Italics means result rejected

Tabel 7

Summary of Results at 95% Confidence Limit

	ALTCP Data A	ALTCP Data B	ALTCP Data C	ALTCP Data D	ALTCP Data E	ALTCP Data F	Total
Number of labs	16	16	22	22	21	19	16
Number of tests methods	13	10	15	16	14	14	82
Number of lab results	198	141	295	255	276	234	1399
Number of last methods with rejection of results:							
T-test	10	6	5	10	6	9	46
b_2-test	10	6	5	10	6	10	47
r-test	9	7	6	11	6	9	46
Number of lab results rejected:							
T-test	16	13	11	15	16	21	92
b_2-test	13	12	15	16	16	23	95
r-test	13	15	13	18	11	21	91
Number of test methods with agreement in results:							
T-test/b_2-test	11	9	13	14	14	12	73
T-test/*r*-test	10	7	12	14	11	11	64
b_2-test/*r*-test	11	7	13	15	11	12	69
all three tests	10	7	12	14	11	11	64
Number of test methods with disagreement in results	3	3	3	2	3	3	17
Number of lab results with conflicting verdicts:	4	6	8	2	8	8	36

Tabel 8

Summary of Results in Rejection of Outliers

	ALTCP Data A	ALTCP Data B	ALTCP Data C	ALTCP Data D	ALTCP Data E	ALTCP Data F	Total
% of test methods with rejection of results:							
95% confidence limit							
T-test	76,92	60,00	33,33	62,50	42,86	64,29	56,10
b_2-test	76,92	70,00	40,00	68,75	35,71	64,29	57,32
r-test	69,93	70,00	40,00	68,75	33,71	64,29	57,32
99% confidence limit							
T-test	76,92	60,00	20,00	56,25	28,57	35,71	45,12
b_2-test	61,54	60,00	20,00	50,00	21,43	42,86	41,46
r-test	69,23	60,00	20,00	50,00	21,43	42,86	42,68
% of lab. results rejected:							
95% confidnece limit							
T-test	8,08	9,22	3,73	5,88	5,80	8,97	6,58
b_2-test	6,57	8,51	5,08	6,27	5,80	9,83	6,79
r-test	6,57	10,64	4,41	7,06	3,99	8,97	6,50
99% confidence limit							
T-test	6,57	8,51	1,36	5,10	2,90	3,85	4,22
b_2-test	5,05	8,51	2,71	4,71	2,54	4,70	4,29
r-test	6,57	7,80	3,05	4,71	1,81	5,98	4,57
% of lab. results with conflicting verdict:							
95% confidence limit	2,02	4,26	2,71	0,08	2,90	2,42	2,57
99% confidence limit	1,52	1,42	2,03	0,04	1,09	2,99	1,57