JOURNAL OF ENGINEERING
annals OF
FACULTY OF ENGINEERING HUNEDOARA
TOME VI (year 2008), FASCICULE 3, (ISSN 1584 – 2673)

YOUDEN PLOT IN DIMENSIONS MEASUREMENT

Jozef PETRÍK, Vojtech MIKLOŠ

TECHNICAL UNIVERSITY OF KOŠICE, FACULTY OF METALLURGY,
DEPT. OF INTEGRATED MANAGEMENT, SLOVAKIA

ABSTRACT:
The capability of the dimensions measurement by ten appraisers of ten samples for Charpy V-notch pendulum impact test was evaluated by Youden plot, analysis of measurement uncertainties, Measurement Systems Analysis (MSA), %P/T, Z-score, Mandel's statistics and t-test. The group of "better" appraisers given by Youden plot was validated by majority of other methods.

KEYWORDS:
Dimension, measurement, Youden plot, capability

1. INTRODUCTIONS

Youden plot (analysis) is directed toward interlaboratory comparisons. Youden's main objective was to determine the precision of a procedure and expect all laboratories to meet this level of precision. For the original Youden plot two samples must be similar and reasonably close in the magnitude of the property evaluated. The axes in this plot are drawn on the same scale: one unit on the x-axis has the same length as one unit on the y-axis. Each point in the plot corresponds to the results of one laboratory (appraiser) and is defined by a first response variable on the horizontal axis and a second response variable on the vertical axis. A horizontal median line is drawn parallel to the x-axis so that there are as many points above the line as there are below it. A second median line is drawn parallel to the y-axis so that there are as many points on the left as there are on the right of this line. The intersection of the two median lines is called the Manhattan median. A circle is drawn that should include 95 % of the laboratories (appraisers) if individual constant errors could be eliminated. A 45-degree reference line is drawn through the Manhattan median. The advantage of using Youden plot is its unique ability to separate random and systematic errors. An Error that is purely systematic will fall on the 45 degree line. A horizontal line drawn from the "45 degree intercept point" to the error vector shows the proper random and systematic components [1].

2. EXPERIMENTAL

Ten appraisers (A, B…J) measured ten randomly selected samples for Charpy V-notch pendulum impact test. The dimensions A (height of sample, 10 mm ± 0.06 mm) and B (width of sample - parallel with the notch, 10 mm ± 0.11 mm according to standard [2]) were measured three times (trials) on each sample and process capability indices C_p and C_{pk} [3, p. 19][4, p. 93, 107] of samples production process were calculated.

Digital micrometer "Kinex" (according to DIN 863) was used as the measurement equipment. It was calibrated at V/2008. The expanded uncertainty of U = 0.001 mm (coverage factor k = 2) and the bias for 12.9 mm standard is 0 mm. The discrimination of the equipment is 0.001 mm. The standard uncertainty u_B = 0.000578 mm. The uncertainty of equipments was calculated according to [5].

As can be seen in fig. 1, all indices C_p and C_{pk} of samples are above 1.33 – the process of their production appears capable. The average dimensions, measured by individual appraisers are on individual samples are at fig. 2.

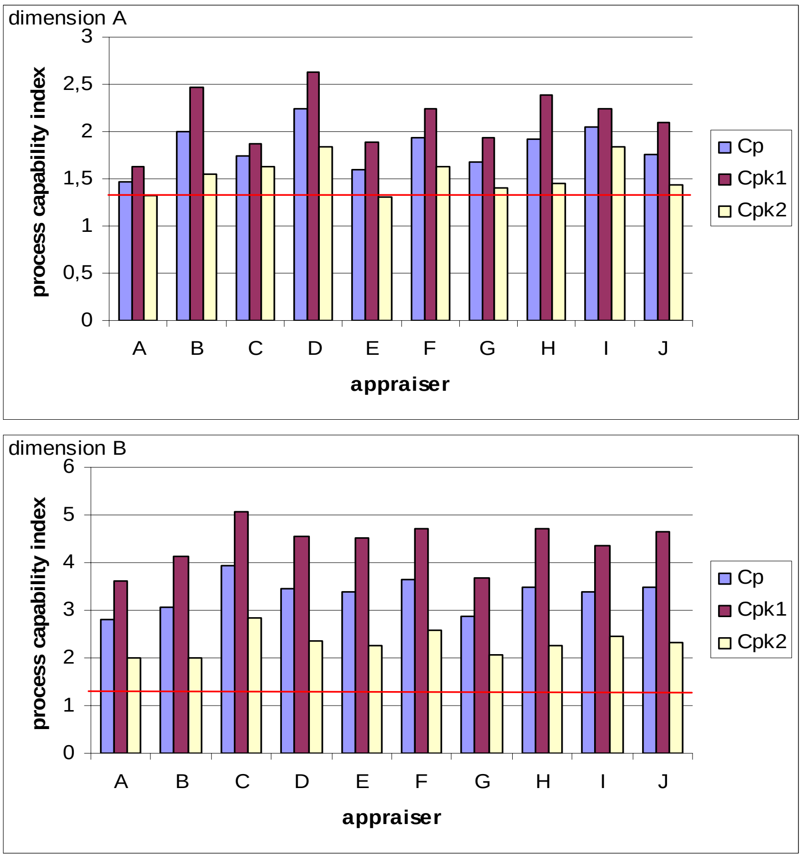


FIGURE 1. IDICES C_p, C_{pk}.

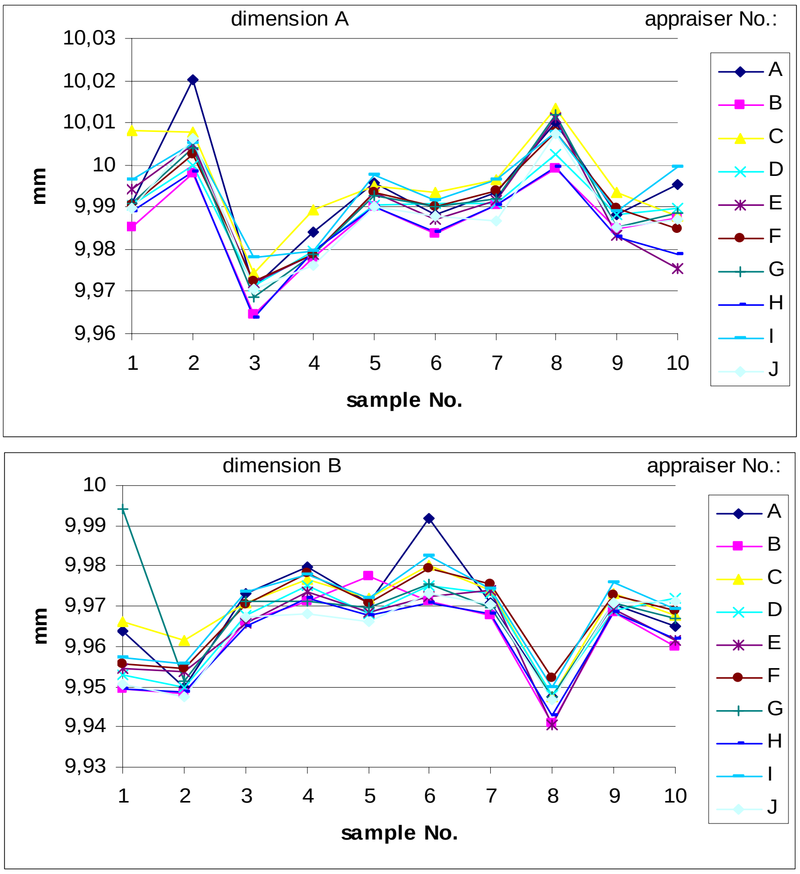


FIGURE 2. AVERAGE DIMENSIONS

The aim of submitted work is to use some analytic methods (computational and graphical) for evaluation of the dimensions measurement process quality and including study of appraisers influence. The Youden plot was used as a start method; the results were verified using MSA, %P/T, Z-score, Mandel's statistics and t-test.

3. PRIMARY STATISTICAL TESTS

The outliers, detected by Grubbs' test (with significant level a = 0,05) were determined for appraisers A, B, C, D, E, H and I. The statistical outliers would indicated, that the process is suffering from special disturbances and is out of statistical control. The normality was estimated by Freeware Process Capability Calculator software, using Anderson – Darling test. The normal distribution have files (the results of all appraisers, measuring one sample) of samples No. 3, 5, 9 10 for dimension A and of samples 3, 4, 7, 8 and 10 for dimension B.

According to Two Factor ANOVA without replication, the influence of the difference between samples and appraisers are statistically significant for the dimension A (height) and dimension B (width) of samples [6, p. 129].

4. UNCERTAINTY

The standard uncertainty of measurement $u_c = \sqrt{u_A^2 + u_B^2}$ (1)

$$u_A = \frac{s}{\sqrt{n}} \quad (2)$$

Where s is standard deviation of trials, measured on one dimension of one sample by one appraiser, the number of the trials n = 3, u_B is thereinbefore uncertainty of used equipment and expanded uncertainty U = k.u_C, with coverage factor k = 2.

The relative expanded uncertainty $U_{rel} = \frac{U}{X}(\%)$ (3)

Where X is relevant average dimension.

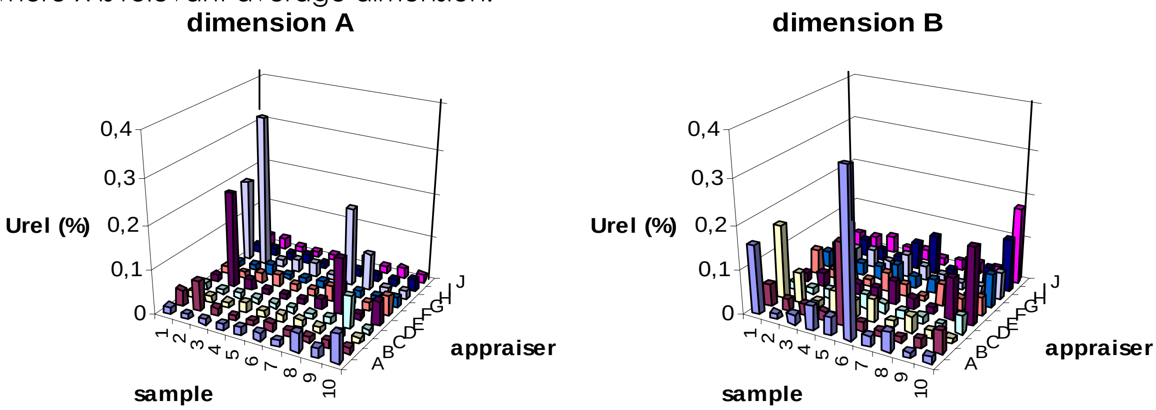


FIGURE 3. THE UNCERTAINTY

As can be seen from fig. 3, higher values of relative expanded uncertainty U_{rel} of individual appraisers were calculated for appraisers E (0.22 %) and H (0.34 %, 0.18 %) measuring the height and for appraisers A (0.37 %), B (0.19 %), C (0.17 %) and J (0.17 %), measuring the width of the samples.

5. YOUDEN PLOT

Youden plot (for average dimensions of all 10 samples) is on the fig. 4. The best results – minimum total error has appraiser C. The values of total error, systematic error and random error are on the fig. 5. Potential source of the systematic error is variability in compressive force of micrometer – a result of incorrect using of equipment by some appraisers (they not used ratchet). According to results of Youden plot – total error – the appraisers were divided into two groups: 1 "better": appraisers D, F, G, E, J and 2 "worse": appraisers A, I, B, H, C.

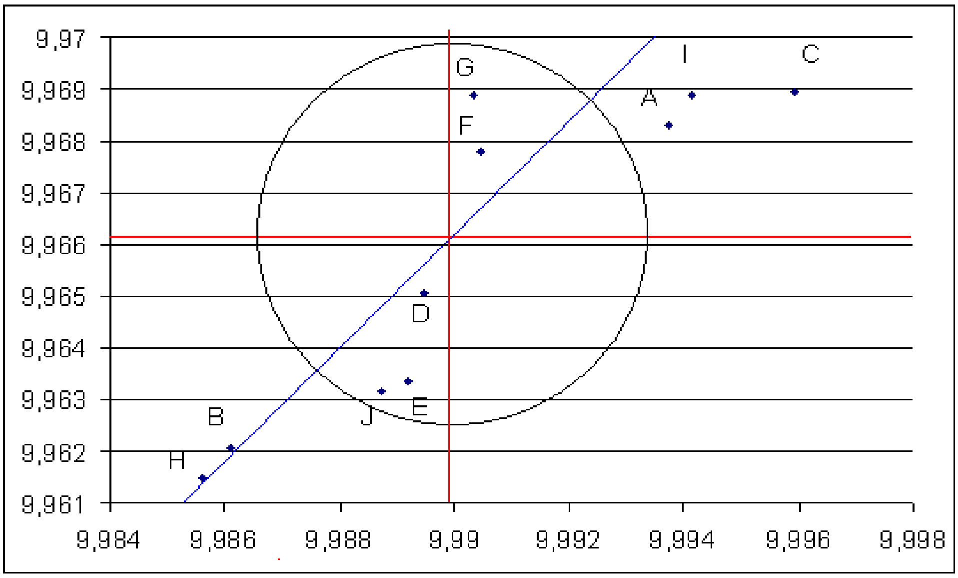


FIGURE 4. YOUDEN PLOT

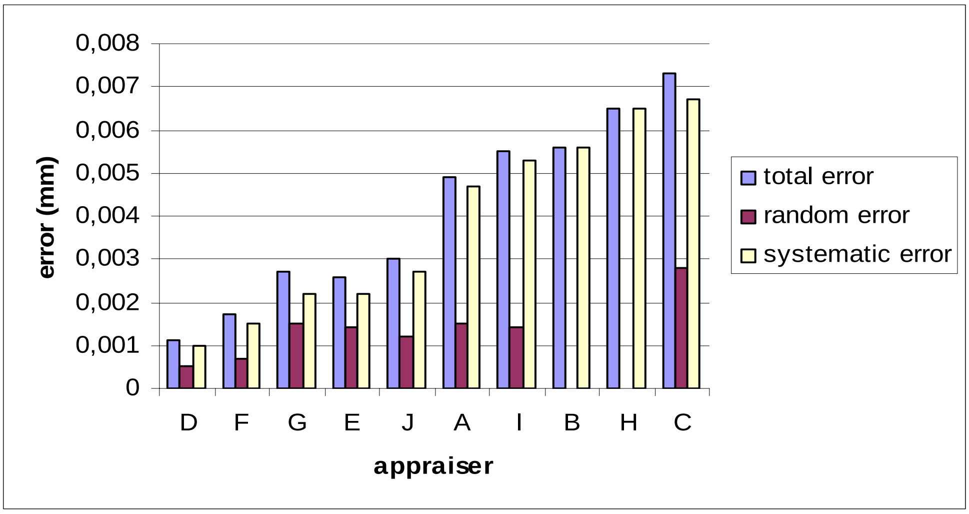


FIGURE 5. TOTAL, SYSTEMATIC AND RANDOM ERROR

6. MEASUREMENT SYSTEMS ANALYSIS (MSA)

The GRR technique of MSA method - combined estimate of measurement system repeatability and reproducibility, described in [3, p. 99-117][7] with confidence level 99 % (5.15 σ) was used for capability evaluation for each for these for groups. The software Palstat CAQ was used for calculation.

The first step of analysis is to estimate whether the discrimination - the value of smallest scale division (graduation) of measurement equipment is sufficient. A general rule of thumb is the discrimination (0.001 mm) ought to be at least one - tenth the process variation. Looking at table 1, we can see, that digital micrometer roughly fulfils this condition [3, p. 74].

Table 1.: The average dimension $\bar{x}$ and standard deviation s and X values out of control limits.

group	dimension of sample	$\bar{x}$ (mm)	s (mm)	X values out of control limits
1	height (A)	9.98964	0.011356	42 %
	width (B)	9.965653	0.01129	30 %
2	height (A)	9.9911	0.012934	34 %
	width (B)	9.96594	0.012478	42 %

The measurement system ought to be in statistical control before capability is assessed, the range control chart is used. The process is in the control, if all ranges are between control limits. As can be seen from table 2, this condition is satisfied for appraisers D, F, G, H, I, the most points out of control limits has appraiser A. If one appraiser is out of control, the method used differs from the others.

Table 2.: The points out of control limits of the range control chart.

appraiser		A	B	C	D	E	F	G	H	I	J
dimension	A	3	-	1	-	-	-	-	-	-	3
	B	2	1	1	-	3	-	-	-	-	1

The area within the control limits of the X-bar control chart represents measurement sensitivity („noise"). Since measurements used in the study represents the process variation, approximately one half or more of the averages should fall outside the control limits. If less than half fall outside the control limits then either the measurement system lacks adequate effective resolution (discrimination) [3, p. 102]. As can be seen in table 1, the condition of sensitivity was not satisfied.

Table 3.: The capability indices.

group	dimension of sample	%EV	%AV	%PV	%GRR	ndc
1	height (A)	35.8	0.0	93.3	35.8	3.675
	width (B)	34.1	22.8	91.2	41.0	3.134
2	height (A)	44.5	30.2	84.3	53.8	2.208
	width (B)	40.3	23.7	88.4	46.7	2.670

The number of distinct categories ("ndc", based on Wheeler's discrimination ratio) is connected with the question of the resolution of measurement equipment. It indicates the number of various categories, which can be distinguished by the measurement systems. It is the number of non-overlay 97 % confidence intervals, which cover the range of expected variability of product. The "ndc" is greater than or equal to 5 for capable processes, results with "ndc" values between 2-5 may be conditionally used for rough estimations [3, p. 117].

The %EV index represents the cumulative influence of measurement equipment, used measuring method and environmental conditions on the variability. It is a function of average range of trials of all appraisers. Whereas the calibration interval of measurement equipment is valid (not expired) and the environmental conditions were stable, worse application of measurement method and possible worse handling with equipment by appraisers of group 2 increased %EV index.

The %AV index represents the influence of appraisers on the variability, for example their responsibility and competence. It is a function of the maximum average appraiser difference. The results of group 1 are better.

The %GRR index refers to contribution of measurement instrument to the variability. Its value represents the process capability in practice. For %GRR<10 % is measurement system capable and for %GRR > 30 % is considered to be not capable (acceptable). As can be seen in table 3, the capability of group 1 is better than capability of group 2 for both dimensions.

All values of %GRR are above 30 % - analyzed process of measurement is not capable, eventually in compliance with "ndc" they can be conditionally used for rough estimations. For example the value of %GRR of measurement process of ten nails □ 4 mm ± 0.11 mm with micrometer was between 34.4 % and 39.7 % [8]. Worse capability has the hardness measuremet process – for Vickers tests %GRR = 66,4 % [9], for Brinell test was %GRR between 62.7 and 88.7 % (wrought brass) [10] or between 63.7 and 89.4 % (cast brass) [11]. Low capability has also process of blood pressure measurement (%GRR between 35 % and 75 % [12][13].

The %PV is a function of the range of individual samples. It is sensitive to influence of variability between dimensions variation. The values of %PV indirectly describe suitability of used measurement equipment for specific measurement: above 90 % are for suitable (group 1) and above 70 % for satisfactory (group 2) [14, p. 29].

7. Z-SCORE

Z-score method, routinely applied in interlaboratory comparison tests was used for validation of above mentioned results. The value for individual sample is:

$$z_i = \frac{x_i - \bar{x}}{s} \tag{4}$$

x_i is the average diameter of all trials on one sample, measured by one appraiser with one micrometer, $\bar{x}$ is height or width of the sample according to standard [2] and „s" is standard deviation of all of one dimension by one micrometer. The results | z_i | ≤ 2 are satisfactory and | z_i | ≥ 3 are unsatisfactory [15, p. 32]. As can be seen on the fig. 6, dimension B (width) of all samples is underestimated without satisfactoriness for all appraisers. As far as the dimension A (height), Z-score for all appraisers is up to 3. The "critical" are the height samples No. 2 and 8 with opposite orientation of Z-score than it of other samples.

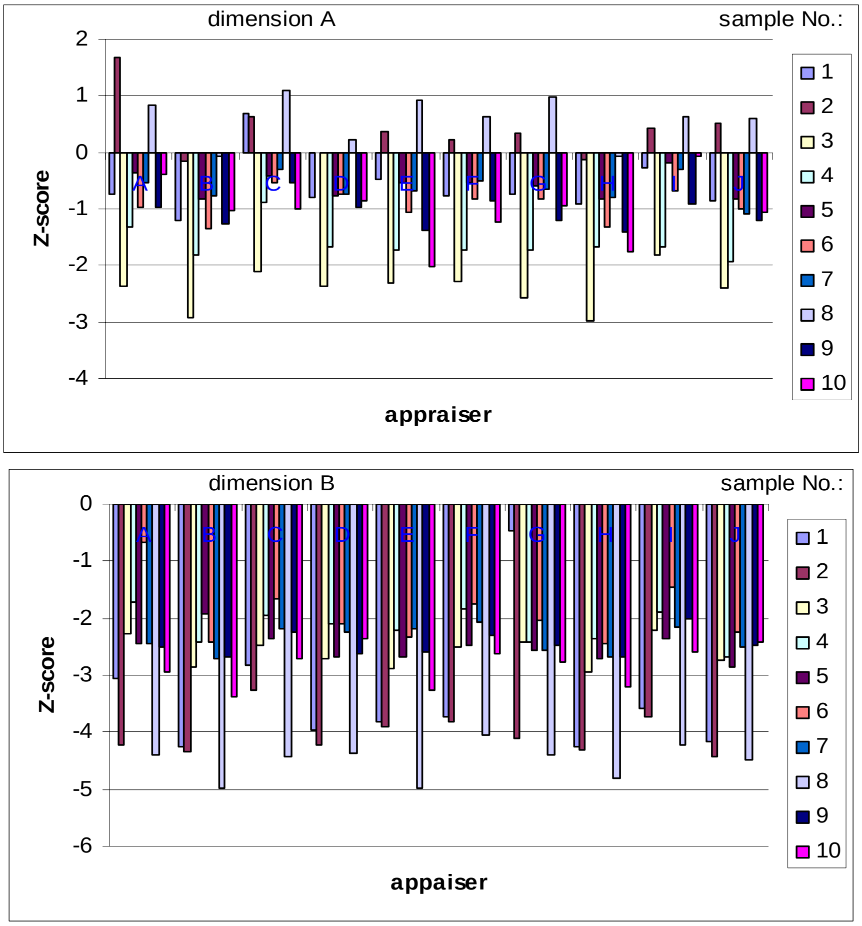


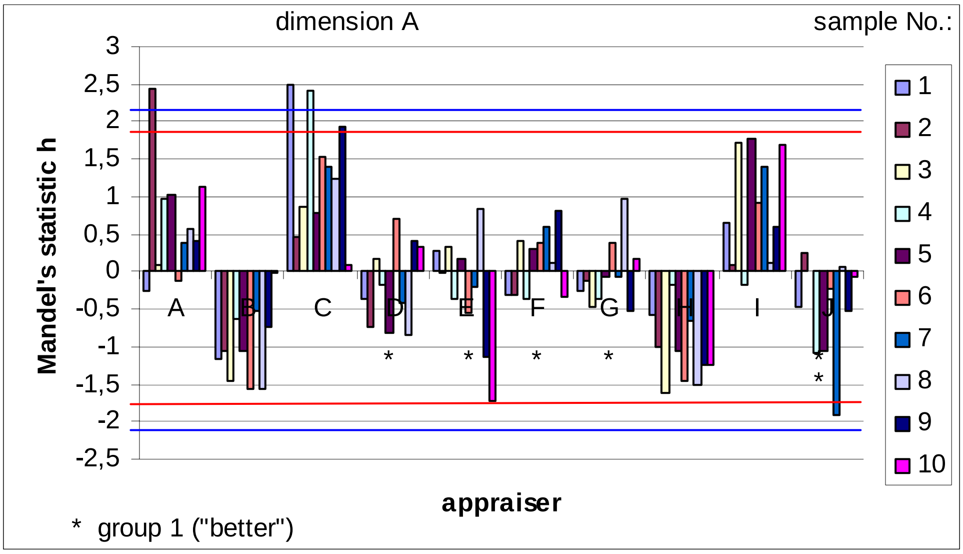
FIGURE 6. Z-SCORE

8. MANDEL'S STATISTICS

Mandel's statistic h is based on average, statistic k is based on variance:

$$h_i = \frac{\bar{x}_i - \bar{\bar{x}}}{\sqrt{\frac{1}{p-1}\sum_{i=1}^{p}\left(\bar{x}_i - \bar{\bar{x}}\right)^2}} \qquad k_i = \frac{s_i\sqrt{p}}{\sqrt{\sum_{i=1}^{p} s_i^2}} \tag{5, 6}$$

Where: $\bar{x}_i$ is average value of one sample measured by one appraiser and $\bar{\bar{x}}$ is average value of all samples measured by all appraisers, s_i is standard deviation of measurements on one sample by one appraiser, p is a number of appraisers. Statistic k gains only positive values. The lines are drawn into the Mandela's statistics plots, representing critical values on the significance level 5 % (strugglers) and 1 % (outliers) [16, p. 217]. As follows from fig. 7 (statistic "h") and fig. 8 (statistic "k"), according to statistic "h" one outlier was obtained for group 1 (appraiser J) and two (appraisers A, C) for group 2 (dimension A) one outlier for group 1 (appraiser G) and two (appraisers A, B) for group 2 (dimension B). According to statistic "k" three outlier were obtained for group 1 (appraisers D, J) and six (appraisers A, C, I) for group 2 (dimension A), six outliers were obtained for group 1 (appraisers F, G, H, J) and eight (appraisers A, B, C, I) for group 2 (dimension B). Less number of outliers is in group 1.



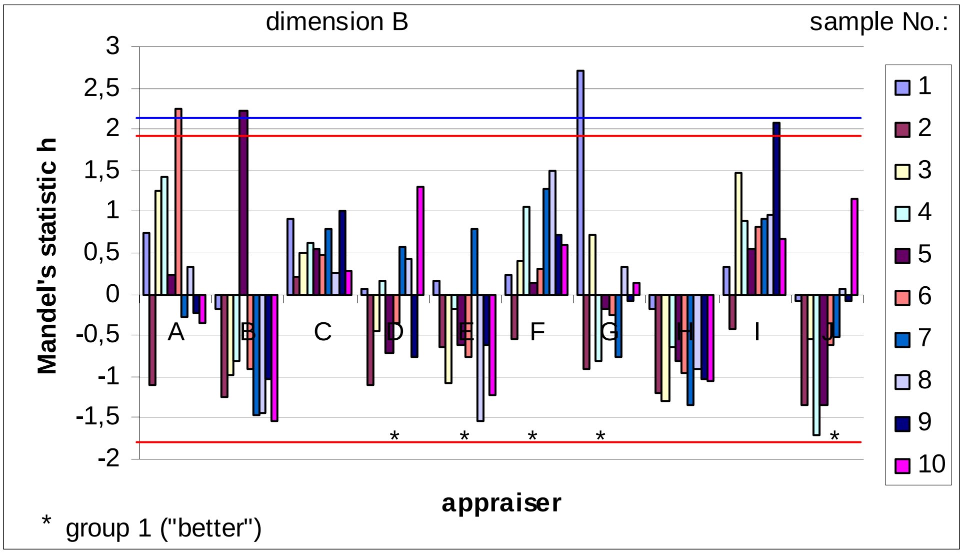


FIGURE 7. MANDEL'S STATISTIC „h".

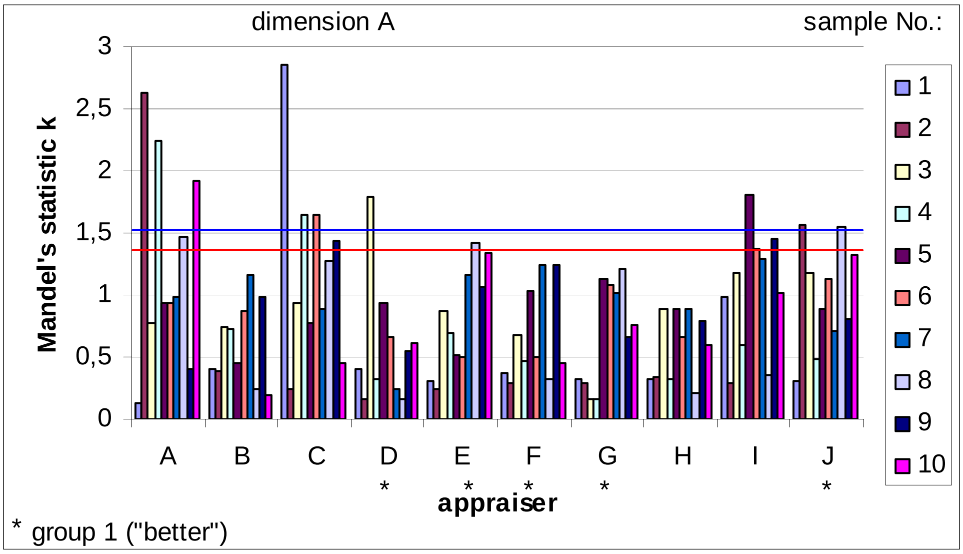
9. AUTOMATED MEASUREMENT TEST EQUIPMENT %P/T

Freeware GRR Calculator V2.0 from Symphony Technologies was used for calculation of automated measurement test equipment. In case of automated test equipment, there is no appraiser influence. The study uses only one operator. It is recommended that 10 parts that span the entire range of the process output should be selected. Each part was measured 3 times (3 trials) by one operator.

$$\%P/T = \frac{6 \times \dfrac{sBar}{c4}}{USL - LSL} \times 100 \qquad (7)$$

Where: sBar = the sum of the standard deviations for each part measured divided by the number of parts in the study, USL, LSL are tolerances of the height and width according to standard [2], c4 is correction factor, depending on the number of trials and used coverage factor = 6σ.

Acceptance criteria for %P/T are the same as for %GRR: under 10 % acceptable, between 10 and 30 % marginal and over 30 % not acceptable. As can be seen from fig. 9, the values of %P/T approximately copy the division into groups according to Youden plot (16.07 % for group 1 and 21.23 % for group 2). The most significant values obtained appraises A, C and E.



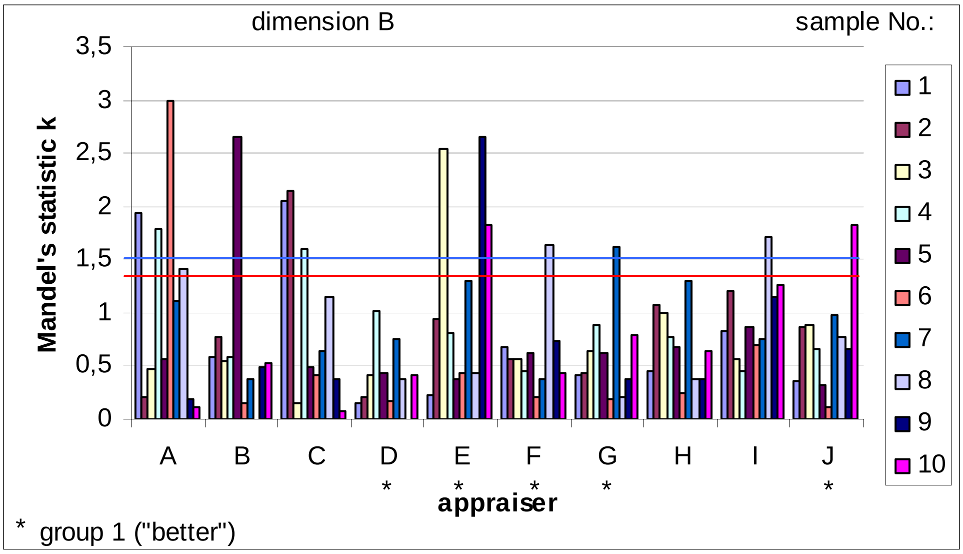


FIGURE 8. MANDEL'S STATISTIC „k“.

10. THE UNPAIRED T-TEST TO COMPARE TWO MEANS

The average (mean) values of the dimension A (height) of all samples measured by individual appraisers were compared by unpaired t-test with 95 % confidence interval. The results are on the fig. 10. The differences between the appraisers within the frame of group usually were not statistically significant in contrast to differences between the appraisers appurtenant to the "better" or "worse" group.

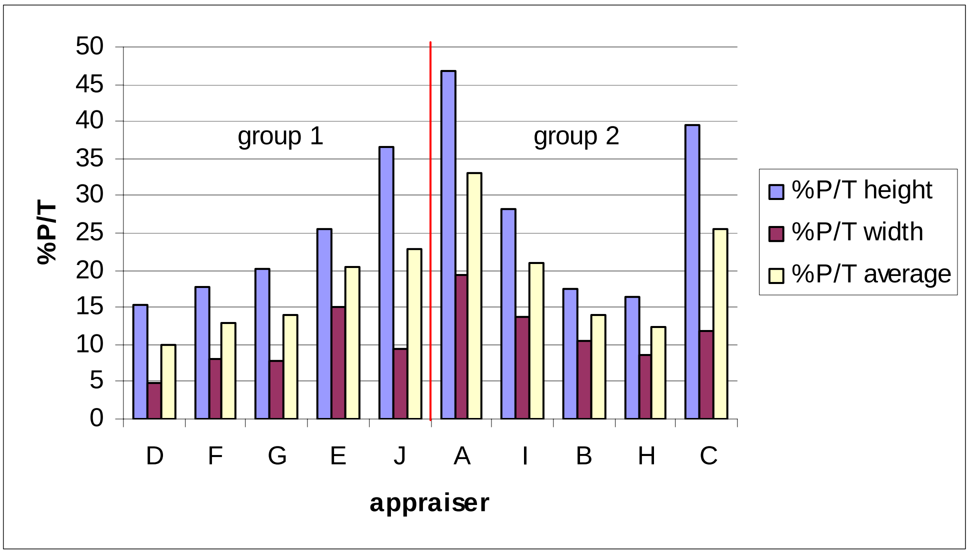


FIGURE 9. THE VALUES OF %P/T

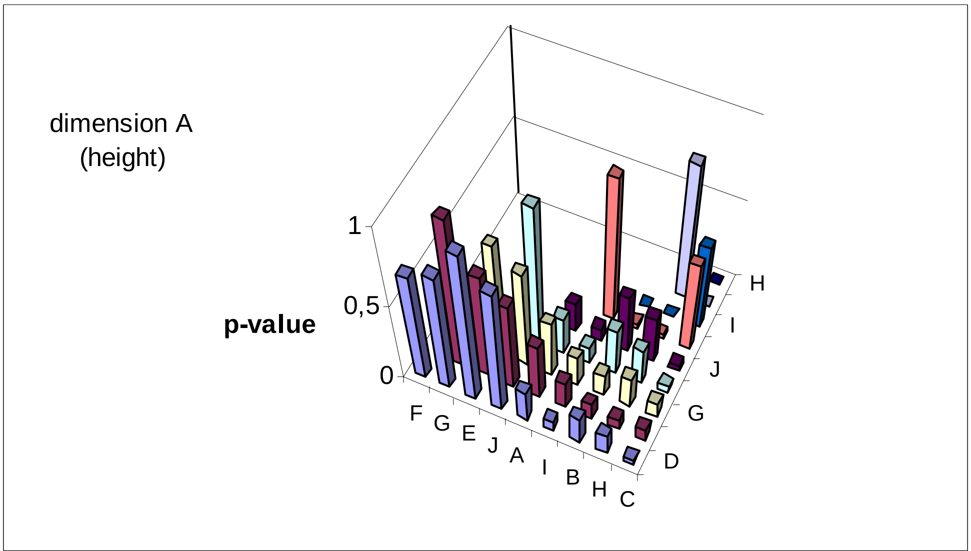


FIGURE 10. P-VALUES

11. CONCLUSION

- Analyzed processes are not capable.
- The insufficient resolution of micrometer are one of the reasons of non – capability, another is low quality (competence) of some appraiser.
- According the results of Youden plot the appraisers were divided into two groups (1 “better” and 2 “worse”), the appraisers D, E, F, G, J are in “better” group.
- MSA verified as a more capable (%AV and %GRR indices) “better” group of appraisers, selected by Youden plot.
- The results of Z-score do not copy the distribution of the appraisers into groups following the Youden plot. The results of the measurement of the height of the samples are more satisfy than those (underestimated) of the width.
- According to Mandela’s statistic “h” group 1 has 2 and group 1 has 4 outliers and according to Mandela’s statistic “k” group 1 has 4 and group 1 has 14 outliers.
- The values of %P/T copies the division into groups according to Youden plot.
- According the t-test to compare two means differences between the appraisers of different groups are more significant than the difference of appraisers within group.
- Regarding the results of all used methods, the appraisers A, B and C are the least capable, appraiser D is the best.

ACKNOWLEDGEMENTS
This work was supported by the Slovak Grant Agency for Science VEGA 1/4141/07

REFERENCES / BIBLIOGRAPHY

[1] [online]. [cited 23 September 2008]. Available from: <http://www.medcalc.be/manual/youdenplot.php>.
[2] STN EN 10045-1:1997. Charpy impact test on metallic materials. Test method (V- and U-notches).
[3] Measurement systems analysis (MSA). Reference manual. Third edition. (2003).
[4] Terek, M. – Hrnčiarová, Ľ.: Analýza spôsobilosti procesu. Bratislava 2001. ISBN 80-225-1443-8.
[5] E. Dietrich, It can be easy. Using the Process Steps of a Measurement System Capability Study to Determine Uncertainty of Measurement. [online]. [cited 23 January 2007]. Available from: < http://www.q-das.de/homepage_e/es%20geht%20auch%20einfach_e.htm>.
[6] Chajdiak, J.: Štatistika v exceli. Bratislava 2002. ISBN 80-85659-27-1.
[7] [online]. [cited 11 November 2005]. Available from:**<**http://jakost.cz/metodiky/download/srovnani-msa-vda.pdf**>.**
[8] Petrík, M.: Capability of dimensions measurement. Journal of Engineering. Annals Faculty of Engineering Hunedoara. Vol. VI (2008), No. 2. ISSN 1584-2665, p. 71-76.
[9] Petrík, J. – Szarvasy, P. – Špeťuch, V.: Hodnotenie merania tvrdosti silumínu s obsahom niklu. Transfer 2004, Trenčín,ISBN 80-8075-030-0, p. 401-406.
[10] Petrík, J. - Špeťuch, V.: Časová závislosť spôsobilosti merania tvrdosti, Transfer 2005, Trenčín, ISBN 80-8075-070-X, p. 429-432.
[11] Petrík, J. - Špeťuch, V.: The dependence of brass hardness measurement system on time. Metrology and measurement systems. No. 4, Vol. 14, (2007), ISSN 0860-8229, p. 595 – 602.
[12] Petrík, J. – Petríková, E.: Spôsobilosť merania krvného tlaku. Laboratórna diagnostika. No. 1, Vol. 12 (2007), ISSN 1335-2644, p. 21-24.
[13] Petrík, J. - Mikloš, V.: Spôsobilosť procesu merania tlaku krvi. Metrológia a skúšobníctvo. No. 1, Vol. 12 (2007), ISSN 1335-2768, p. 15-19.
[14] Meloun, M. – Militký, J.: Kompendium statistického zpracování dat. Praha 2002. ISBN: 80-200-1008-4.
[15] MSA 0113-98. Methodical guidelines for accreditation. Proficiency testing. Bratislava 1998.
[16] Palenčar, R. et al.: Štatistické metódy v metrologických a skúšobných laboratóriách. Bratislava 2001, ISBN 80-968449-3-8.