# BLUNDER DETECTION IN GEODETIC NETWORKS

*Elena Maria POPA, PhD student – Technical University of Civil Engineering Bucharest, elenamaria.popa@gmail.com*
*Ciprian Bogdan SPĂTAR, PhD student – Newcastle University , c.b.spatar@ncl.ac.uk*

***Abstract**: In statistics, observations with numerical values which differ significantly from the other measurements are denoted as outliers. These values must be identified and removed from the set of observations before performing a least squares adjustment, which would otherwise yield poor or invalid results. Unidentified blunders in the preprocessing stage (a priori) can be isolated after the adjustment (a posteriori) using tailored statistical tests. This paper reviews procedures for identifying outlying observations before and after conducting a geodetic network adjustment.*

***Keywords:** blunder, systematic error, random error, least squares adjustment*

## 1. Introduction

Measurements are affected by errors classified as blunders, systematic errors and random errors. Out of these three categories, only random errors constitute the subject of the theory of errors and observation processing.

Blunders are unusually large size errors caused by operator, number transposition, entry and recording errors, serious malfunction of the instruments or the use of inappropriate measurement methods. It is desirable to remove blunders from the data set, no matter their sizes.

Before the procedures for identifying blunders presented in the following paragraphs is advisable to check normality observations through statistical tests for this purpose.

## 2. A priori methods for detecting blunders

2.1 Checking the random nature of observations errors

Random errors have the following characteristics:

1. The arithmetic mean observations error $\varepsilon_i$ $\left(i = 1, 2, \dots, n\right)$ approaches zero when the number of observations $n$ tends to infinity:

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} \varepsilon_i}{n} = 0; \tag{1}$$

2. Positive and negative errors with the same absolute value have equal chance of occurrence;
3. Small absolute errors occur more often than with large absolute values.
4. Error values must fall within certain limits.

Statistical tests can be built to verify the random error of the defining features of statistical tests to identify non-random errors are (Fan, 1997):

- Null hypothesis $H_0$ is the assumption that measurement errors are random and data distributions. Alternative hypothesis argues that measurement errors are not considered part of the distribution
- Test statistics can be calculated using observations and a well defined distribution;
- For each observation we compare statistics with the critical value at a significance level. If the test passes are acceptable, and if it can not be suspected of systematic error or blunders.

2.2 Grubbs test

Grubbs test was proposed by Frank E. Grubbs in his work *Procedures for detecting Outlying Observations in Samples*, published in 1969. This test is used to identify external values in a series of observations $x_1, x_2, \ldots, x_n$ normally distributed. It is advisable to check data normality before applying this test.

Steps to Grubbs test are:

1. Calculate the average selection using next relation

$$\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i; \tag{2}$$

2. Calculate the standard deviation of selection using the expression

$$s_0 = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} \left( x_i - \overline{x} \right)^2}; \tag{3}$$

3. Calculate statistics

$$G = \frac{\left| x_e - \overline{x} \right|}{s_0}, \tag{4}$$

where $x_e$ is a selection of extreme values (minimum or maximum value)

4. Extract from Table 1 the critical value $G(n,\alpha)$ depending on the amount of selection $n$ and materiality $\alpha$;
5. Observation is removed if

$$G > G(n,\alpha); \tag{5}$$

6. Repeat steps 1-5 to eliminate all erroneous observations.

Grubbs test is not recommended for selection volume less than 7 because in this case most of the observations will be labeled as external values. Also, for a volume greater than 25 selection results are obtained with a rough approximation.

Table 1. Critical values for the Grubbs test

| ***n*** | $\alpha$ | | ***n*** | $\alpha$ | | ***n*** | $\alpha$ | |
|---|---|---|---|---|---|---|---|---|
| | ***0.05*** | ***0.01*** | | ***0.05*** | ***0.01*** | | ***0.05*** | ***0.01*** |
| 3 | 1.1531 | 1.1546 | 15 | 2.4090 | 2.7049 | 80 | 3.1319 | 3.5208 |
| 4 | 1.4625 | 1.4925 | 16 | 2.4433 | 2.7470 | 90 | 3.1733 | 3.5632 |
| 5 | 1.6714 | 1.7489 | 17 | 2.4748 | 2.7854 | 100 | 3.2095 | 3.6002 |
| 6 | 1.8221 | 1.9442 | 18 | 2.5040 | 2.8208 | 120 | 3.2706 | 3.6619 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 7 | 1.9381 | 2.0973 | 19 | 2.5312 | 2.8535 | 140 | 3.3208 | 3.7121 |
| 8 | 2.0317 | 2.2208 | 20 | 2.5566 | 2.8838 | 160 | 3.3633 | 3.7542 |
| 9 | 2.1096 | 2.3231 | 25 | 2.6629 | 3.0086 | 180 | 3.4001 | 3.7904 |
| 10 | 2.1761 | 2.4097 | 30 | 2.7451 | 3.1029 | 200 | 3.4324 | 3.8220 |
| 11 | 2.2339 | 2.4843 | 40 | 2.8675 | 3.2395 | 300 | 3.5525 | 3.9385 |
| 12 | 2.2850 | 2.5494 | 50 | 2.9570 | 3.3366 | 400 | 3.6339 | 4.0166 |
| 13 | 2.3305 | 2.6070 | 60 | 3.0269 | 3.4111 | 500 | 3.6952 | 4.0749 |
| 14 | 2.3717 | 2.6585 | 70 | 3.0839 | 3.4710 | 600 | 3.7442 | 4.1214 |

Chauvenet's criterion and Q test (Dixon) can also be used to identify external values in a series of observations made on a single size.

2.3 Use of constant terms vector

Constant term, obtained as difference between the value calculated using provisional coordinates and the measured value of a quantity, can provide indications of the existence of errors, especially if it has a higher numeric value. If the observations are wrong two situations are possible (Ghilan, 2010):

- If a wrong observation is not used to calculate initial coordinates of a point from geodetic network, the corresponding element from constant terms vector has a large value;
- If a wrong observation is used to calculate initial coordinates, the remaining redundant observations should have relatively large values.

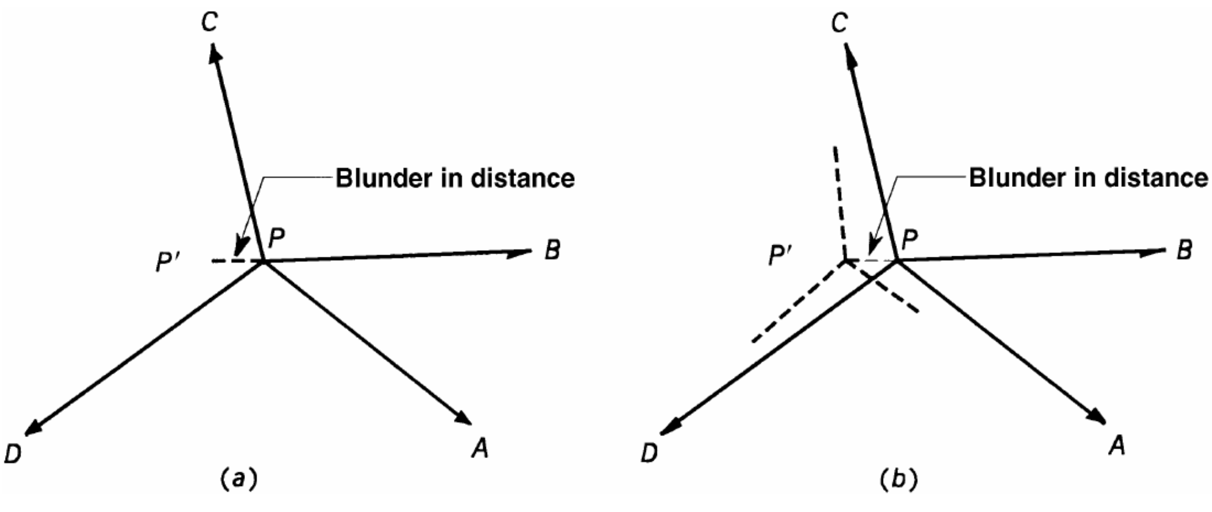


Fig. 1. Influence of blunder on the initial coordinates of points

Figure 1 illustrates the two situations (after Ghilan, 2010). In Figure 1 (a) the wrong observation is the distance $BP$ and blunder is $PP'$. This distance is not used to calculate the initial coordinates of point *P* and therefore the corresponding element of constant term vector will have a large value. In Figure 1 (b) wrong observation is used to calculate initial coordinates of point *P*. Therefore, the redundant angular and linear observations connecting point *P* to points *A*, *C* and *D* have corresponding elements with large values in constant terms vector.

## 3. A posteriori blunder detection

3.1 Data-snooping

This procedure was proposed by Willem Baarda in his work *A testing procedure for use in geodetic networks* in 1968. Data-snooping is a method for identifying wrong observations based on preliminary processed by the method of least squares and a statistical test applied individually to each corrections obtained. It is assumed that only one observation contains a blunder.

Consider a geodetic network in which were made $n$ observations contained in the vector

$$\boldsymbol{m}^* = \left[ m_1^*, m_2^*, \dots, m_n^* \right]^T \tag{6}$$

to determine a number of $u$ parameters

$$\boldsymbol{X} = \left[ X_1, X_2, \dots, X_u \right]^T . \tag{7}$$

The data-snooping null hypothesis $H_0$ is as follows: no observations affected by blunders, and functional-stochastic model of processing by the method of least squares is the equation

$$H_0 : \begin{cases} \underset{(n\times 1)}{\boldsymbol{v}} = \underset{(n\times u)}{\boldsymbol{A}} \underset{(u\times 1)}{\boldsymbol{x}} + \underset{(n\times 1)}{\boldsymbol{l}} \\ \boldsymbol{C}_{mm} = \sigma_0^2 \boldsymbol{Q}_{mm} = \sigma_0^2 \boldsymbol{P}^{-1} , \\ \boldsymbol{v}^T \boldsymbol{P} \boldsymbol{v} \to \text{minin} \end{cases} \tag{8}$$

where **v** is the vector of corrections, **A** the variable coefficients matrix, called the geodetic network configuration matrix, **x** the vector of variables, **l** constant terms vector, $\boldsymbol{C}_{mm}$ variance-covariance matrix of measurements, $\boldsymbol{Q}_{mm}$ measurements cofactors matrix, **P** the weight matrix and $\sigma_0^2$ a scalar called unit weight range or variance factor. Variable vector is determined by the equation

$$\boldsymbol{x} = -\left( \boldsymbol{A}^T \boldsymbol{P} \boldsymbol{A} \right)^{-1} \boldsymbol{A}^T \boldsymbol{P} \boldsymbol{l}. \tag{9}$$

Variables cofactors matrix $\boldsymbol{Q}_{xx}$ and cofactors corrections matrix $\boldsymbol{Q}_{vv}$ are

$$\boldsymbol{Q}_{xx} = \left( \boldsymbol{A}^T \boldsymbol{P} \boldsymbol{A} \right)^{-1} , \tag{10}$$

$$\boldsymbol{Q}_{vv} = \boldsymbol{P}^{-1} - \boldsymbol{A} \left( \boldsymbol{A}^T \boldsymbol{P} \boldsymbol{A} \right)^{-1} \boldsymbol{A}^T = \boldsymbol{P}^{-1} - \boldsymbol{A} \boldsymbol{Q}_{xx} \boldsymbol{A}^T . \tag{11}$$

Sum of squares corrections multiplied by the weight are computed using the expression

$$\Omega = \boldsymbol{v}^T \boldsymbol{P} \boldsymbol{v}. \tag{12}$$

Alternative hypothesis: there is one observation $m_i^*$ which contains one blunder. Stochastic model of compensation is the same as for the null hypothesis, but the blunder $\Delta_i$ is also included in the functional model:

$$H_A : \begin{cases} \underset{(n\times1)}{\boldsymbol{v}} = \underset{(n\times u)}{\boldsymbol{A}}\,\underset{(u\times1)}{\boldsymbol{x}} + \underset{(n\times1)}{\boldsymbol{l}} + \underset{(n\times1)}{\boldsymbol{c}_i}\,\Delta_i \\ \boldsymbol{C}_{mm} = \sigma_0^2 \boldsymbol{Q}_{mm} = \sigma_0^2 \boldsymbol{P}^{-1}, \\ \boldsymbol{v}^T \boldsymbol{P} \boldsymbol{v} \to \text{minin} \end{cases} \tag{13}$$

where $\boldsymbol{c}_i$ is a column vector with mull elements, excepting element *i*, the corresponding observation error, which has unit value:

$$\boldsymbol{c}_i = \left[0,0,\ldots,0,1,0,\ldots,0,0\right]^T. \tag{14}$$

If $\boldsymbol{v}_\Delta$ is the estimator obtained by the least squares method of correction vector v under the hypothesis $H_A$, it can be shown that between the value $\boldsymbol{v}_\Delta^T \boldsymbol{P} \boldsymbol{v}_\Delta$ and sum of squares corrections multiplied by the weight given by expression (12) have the following equation:

$$\Omega_\Delta = \boldsymbol{v}_\Delta^T \boldsymbol{P} \boldsymbol{v}_\Delta = \boldsymbol{v}^T \boldsymbol{P} \boldsymbol{v} - \Delta\Omega = \Omega - \Delta\Omega, \tag{15}$$

where

$$\Delta\Omega = \boldsymbol{v}^T \boldsymbol{P} \boldsymbol{c}_i \left(\boldsymbol{c}_i^T \boldsymbol{P} \boldsymbol{Q}_{vv} \boldsymbol{P} \boldsymbol{c}_i\right)^{-1} \boldsymbol{c}_i^T \boldsymbol{P} \boldsymbol{v}, \tag{16}$$

which is determined by the relation (11).

The blunder $\Delta_i$ is obtained using the expression

$$\Delta_i = \left(\boldsymbol{c}_i^T \boldsymbol{P} \boldsymbol{Q}_{vv} \boldsymbol{P} \boldsymbol{c}_i\right)^{-1} \boldsymbol{c}_i^T \boldsymbol{P} \boldsymbol{v}, \tag{17}$$

and the vector of variable parameters is

$$\boldsymbol{x}_\Delta = \boldsymbol{x} - \left(\boldsymbol{A}^T \boldsymbol{P} \boldsymbol{A}\right)^{-1} \boldsymbol{A}^T \boldsymbol{P} \boldsymbol{c}_i \Delta_i = -\left(\boldsymbol{A}^T \boldsymbol{P} \boldsymbol{A}\right)^{-1} \boldsymbol{A}^T \boldsymbol{P}\left(\boldsymbol{l} + \boldsymbol{c}_i \Delta_i\right), \tag{18}$$

where **x** is the vector of variables obtained under the null hypothesis $H_0$ using equation (9).

In stochastic independent observations case, hypothesis generally accepted in current processing, weight matrix P and cofactors measurements matrix $\boldsymbol{Q}_{mm}$ are diagonal matrices and equations (16) and (17) are simplified as follows:

$$\Delta\Omega = \frac{v_i^2}{q_{v_i v_i}}; \tag{19}$$

$$\Delta_i = \frac{v_i}{p_i \cdot q_{v_i v_i}}, \tag{20}$$

where $i = 1,2,\ldots,n$, $p_i$ is the weight of observation $m_i^*$, $v_i$ correction of the same observation, and $q_{v_i v_i}$ is the weighting coefficient *i* from the diagonal of the matrix $\boldsymbol{Q}_{vv}$ [equation (11)].

If the a priori (before compensation) variance factor's value $\sigma_0^2$ is known, we obtain the following two statistics under the null hypothesis:

$$\frac{\Omega}{\sigma_0^2} \sim \chi^2\left(n-u\right); \tag{21}$$

$$\frac{\Delta\Omega}{\sigma_0^2} \sim \chi^2\left(1\right) \text{ equivalent to } \frac{\sqrt{\Delta\Omega}}{\sigma_0} \sim N\left(0,1\right), \tag{22}$$

where $\chi^2(n-u)$ is the distribution $\chi^2$ with $n-u$ degrees of freedom, and $N(0,1)$ is the standard normal distribution (also called standardized normal distribution).

By introducing the relation (19) into (22) obtain statistics

$$w_i = \frac{\sqrt{\Delta\Omega}}{\sigma_0} = \frac{v_i}{\sigma_0\sqrt{q_{v_iv_i}}} \sim N(0,1). \tag{23}$$

If the value of a priori variance factor $\sigma_0^2$ is not known, we can calculate a posteriori value $s_0^2$ using the relation

$$s_0^2 = \frac{\boldsymbol{v}^T\boldsymbol{P}\boldsymbol{v}}{n-u} = \frac{\Omega}{n-u}. \tag{24}$$

Using the value of equation (23) we obtain the following statistical distribution of $t$ (Student):

$$\overline{v}_i = \frac{v_i}{s_0\sqrt{q_{v_iv_i}}} = \frac{v_i}{s_{v_i}} \sim t(n-u-1), \tag{25}$$

which represent the standard correction. The empirical standard deviation of correction $v_i$ was noted with $s_{v_i}$

Equations (21), (23) and (25) can be used to test observations affected by blunders:

a) Global test (multidimensional)

If

$$\Omega < \sigma_0^2 \cdot \chi_{n-u,\alpha} \tag{26}$$

null hypothesis is accepted, otherwise the null hypothesis is rejected. The quantity $\chi_{n-u,\alpha}$ is the critical value of distribution $\chi^2$ with $n-u$ degrees of freedom for the significance level $\alpha$. Generally, the processing geodetic observations $\alpha$ received one of the following values: 0.05 (5%), 0.01 (1%), 0003 (0.3%) or 0001 (0.1%).

b) Test of each individual observation (one-dimensional):

If

$$|w_i| < n_{\frac{1}{2}\alpha} \tag{27}$$

null hypothesis is accepted, otherwise the null hypothesis is rejected. The quantity $n_{\frac{1}{2}\alpha}$ is standard normal distribution critical value $N(0,1)$ for the significance level $\alpha$.

If the value $\sigma_0^2$ is unknown, we use the statistics computed with equation (25) in the one-dimensional test. If

$$|\overline{v}_i| < t_{\frac{1}{2}\alpha}(n-u-1) \tag{28}$$

null hypothesis is accepted, otherwise the null hypothesis is rejected. The quantity $t_{\frac{1}{2}\alpha}(n-u-1)$ is the critical value of distribution $t$ for degrees of freedom and the significance level $\alpha$.

Expressions (27) and (28) can be written in the following forms:

$$\frac{|v_i|}{\sqrt{q_{v_i v_i}}} < \sigma_0 \cdot n_{\frac{1}{2}\alpha}; \tag{29}$$

$$\frac{|v_i|}{\sqrt{q_{v_i v_i}}} < s_0 \cdot t_{\frac{1}{2}\alpha}(n-u-1), \tag{30}$$

which can also be used to identify erroneous observations.

As any statistical test can distinguish two types of error:

- First order error is rejecting the null hypothesis, although this is actually true. The probability of this error is equal to the significance level;
- Second order error is accepting the null hypothesis, although it is actually false. Probability of second order error is $\beta$. The quantity $1-\beta$ is called the power of the test.

Critical values are extracted from statistical tables based on the significance level and volume selection. Can be used as an alternative critical values calculated by Baarda (1968) for different significance levels, which are presented in Table 2.

Table 2. Critical values for data-snooping

| $\alpha$ | $1-\alpha$ | $\beta$ | $1-\beta$ | Critical value |
|---|---|---|---|---|
| 0.05 | 0.95 | 0.80 | 0.20 | 2.8 |
| 0.001 | 0.999 | 0.80 | 0.20 | 4.1 |
| 0.001 | 0.999 | 0.999 | 0.001 | 6.6 |

The theory of errors, for normal distribution 99.9% by errors (or corrections) are within the range $(-3.29 \cdot s_0, 3.29 \cdot s_0)$. Therefore the value of 3.29 can also be used as critical value.

The presence of blunders in the series of observations is illustrated graphically in Figure 2. Alternative distribution represented on the right side of the figure is usually variable. This may be just a normal distribution, but with a different mean and standard deviation. Continuous vertical lines represent the critical value, delimiting critical regions where the null hypothesis is rejected

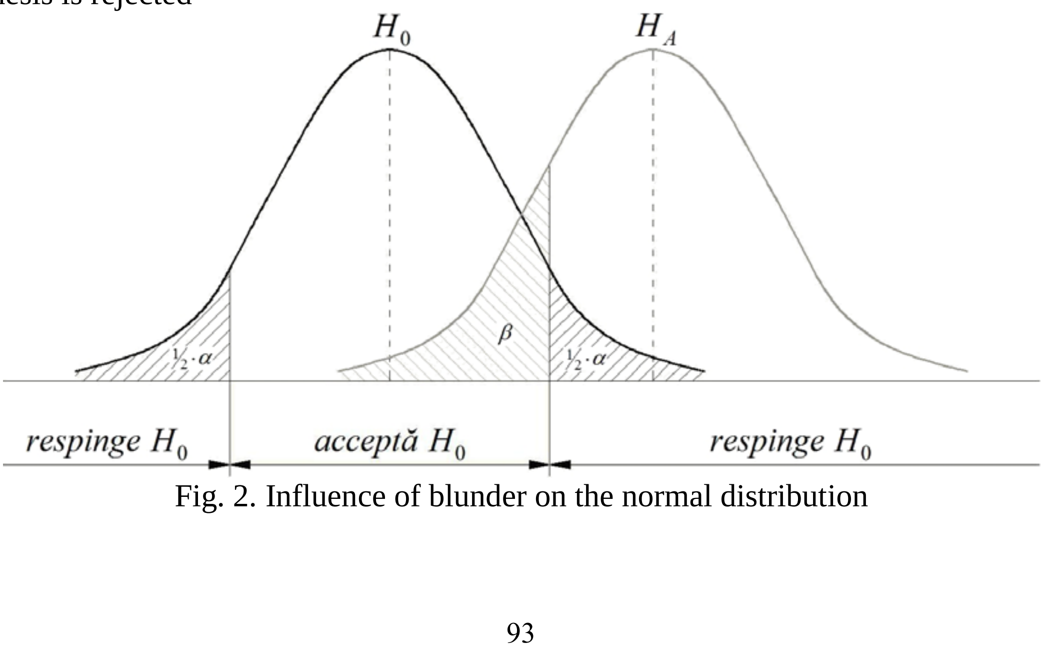


Fig. 2. Influence of blunder on the normal distribution

The steps to be taken for identify and eliminate erroneous observations using data-snooping method are:

1. Identify observations suspected to be affected by errors using one of the expressions (27), (28), (29) or (30);
2. Identify errors causes and eliminate observations with the large absolute value of statistics;
3. Reassume the adjustment;
4. Resumption steps 1-3 until all detected blunders are removed;
5. If more than one observation was removed they can be reintroduced sequentially in the adjustment and tested.

Data-snooping method has certain limitations in identifying blunders. First, the method is sensitive way of estimating the weights geodetic observations, in particular for processing heterogeneous observations. It is useful for this purpose establish an optimal rate of weights between groups of measurements according to known methods in the literature. Also, data-snooping assumes the existence of a single blunder and can not guarantee detection of the presence of several errors. Third, the method of least squares tends to equalize corrections observations across the geodetic network so that it is possible that observations affected by errors can not be precisely identified based on their correction.

3.2 Tau test

This test was developed by Alan J. Pope and presented in the paper *The statistics of residuals and the detection of outliers*, published in 1976. Standard correction obtained with equation (25) is considered by Pope a statistical distribution $\tau$ (tau)

$$\overline{v}_i = \frac{v_i}{s_0\sqrt{q_{v_i v_i}}} = \frac{v_i}{s_{v_i}} \sim \tau\left(n-u\right), \tag{31}$$

where $\tau\left(n-u\right)$ is the distribution $\tau$ with $n-u$ degrees of freedom.

Null hypothesis $H_0$ and alternative hypothesis $H_A$ of this test are identical to those of data-snooping. If

$$\frac{\left|v_i\right|}{\sqrt{q_{v_i v_i}}} < s_0 \cdot \tau_{\frac{1}{2}\alpha}\left(n-u\right) \tag{32}$$

null hypothesis is accepted, that is no errors. Otherwise the null hypothesis is rejected. Term $\tau_{\frac{1}{2}\alpha}\left(n-u\right)$ is the critical value distribution $\tau$ with $n-u$ degrees of freedom for the significance level and can be determinate with the relation

$$\tau_{\frac{1}{2}\alpha}\left(n-u\right) = \frac{\sqrt{n-u} \cdot t_{\frac{1}{2}\alpha}\left(n-u-1\right)}{\sqrt{n-u-1+t^2_{\frac{1}{2}\alpha}\left(n-u-1\right)}}. \tag{33}$$

For an infinite number of degrees of freedom the distribution $\tau$ converge to distribution *t* or to standard normal distribution.

Procedures to identify and eliminate blunders using expression (32) is identical to that used in data-snooping.

3.3 Danish method

This method was proposed by Torben Krarup, etc. in 1980 in his *Götterdämmerung over least squares adjustment*. It is an iterative method that starts with a adjustment method of least squares in which all observations have unitary weight. After initial compensation is awarded one new weight each observation by the size correction of observation, in accordance with the relation

$$p_i = \begin{cases} 1 & \text{dacă } v_i \le 2 \cdot s_0 \\ c_1^{-c_2 \cdot v_i^2} & \text{dacă } v_i > 2 \cdot s_0 \end{cases}, \tag{34}$$

where $c_1$ and $c_2$ are two positive numbers chosen empirically.

Resume adjustment using new weights obtained with equation (34). Iterations continue until the blunders have null weight. The sizes of corrections are a dimension of blunders quantities. Method works efficient for geodetic networks with large redundancy.

## 4. Conclusion

The most used methods to identify blunders in geodetic networks are those apply a posterior. Of these there are data-snooping and test $\tau$ that in practice have identical results. Therefore, the option to use one of the procedures is just a user preference.

Data-snooping and test $\tau$ theoretical based on the theoretical principle that the blunders are treated as random errors with very large values. The corrections observations obtained after an preliminary adjustment are statistically tested to verify the presence or absence of errors. If a certain correction exceeds critical value calculated in the statistic test that observation is suspected to be blunder. Removing suspect observations is dependent upon an analysis of the causes that have brought those values, especially because using the least squares method a error affect the corrections values of other observations.

It is indicated that the geodetic network where is intended to identify errors to be processed as free network, so that errors coordinates of the old points do not affect the process for identifying errors.

Besides methods of identifying a posteriori blunders presented above can be mentioned: the simultaneous determination of blunders and variance components, the minimum amount method and use robust estimators. The last two methods are not based on the classical least squares method: minimize sum of squares corrections multiplied by the weight. These methods change this principle for the whole process of adjustment to be less exposed to the presence of errors or for their easier detection.

## 5. References

1. *Baarda, W. 1968. - A testing procedure for use in geodetic networks. Netherlands Geodetic Commission, Publications on geodesy, Volume 2, Number 5, Delft.*
2. *Danciu, V., Rus, T. 2005. - Aplicarea testelor statistice postprocesare. Revista de Cadastru RevCAD'05, Aeternitas, Alba Iulia.*
3. *Fan, H. 1997. - Theory of Errors and Least Squares Adjustment. Division of Geodesy, Royal Institute of Technology (KTH), Stockholm.*
4. *Ghilani, C.D. 2010. - Adjustment computations: spatial data analysis, 5th edition. Wiley, New York.*
5. *Leick, A. 2004. - GPS Satellite Surveying, 3rd edition. Wiley, New York.*

6. *Moldoveanu, C. 2002. - Geodezie: noțiuni de geodezie fizică şi elipsoidală, poziționare. MatrixRom, Bucureşti.*
7. *Teunissen, P.J.G. 2006. - Network quality control. VSSD, Delft.*
8. *Teunissen, P.J.G. 2006. - Testing theory: an introduction, 2nd edition. VSSD, Delft.*