# Residual Analysis and Detection of Outliers in Mixed Linear Models


## Dimitrios A. Rossikopoulos

*Department of Geodesy and Surveying, Aristotle University of Thessaloniki*



**Περίληψη:** Στην εργασία αυτή αναλύεται το πρόβλημα της αξιολόγησης των παρατηρήσεων σ’ ένα μικτό γραμμικό μοντέλο χρησιμοποιώντας τις τεχνικές του στατιστικού ελέγχου των ακραίων τιμών, οι οποίες στην περίπτωση αυτή των μικτών μοντέλων παρουσιάζουν ιδιαίτερες δυσκολίες, καθώς ελέγχονται τα ολικά υπόλοιπα και δεν μπορεί να γίνει ο διαχωρισμός των τυχαίων επιδράσεων από τα σφάλματα των παρατηρήσεων. Αντιμετωπίζεται και η περίπτωση κατά την οποία προκύπτει το πρόβλημα της εμφάνισης αγνώστων συνιστωσών της μεταβλητότητας αναφοράς, ένα θέμα που συνοδεύει συνήθως τα μικτά μοντέλα. Τότε, και όταν οι βαθμοί ελευθερίας είναι μικροί, εμφανίζεται το πρόβλημα υπολογισμού των βαθμών ελευθερίας των κατανομών που χρησιμοποιούνται στους παραπάνω στατιστικούς ελέγχους.



**Λέξεις κλειδιά:** Μικτά γραμμικά μοντέλα, στατιστικοί έλεγχοι υποθέσεων, βαθμοί ελευθερίας, έλεγχος σφαλμάτων.



**Abstract:** In this article the problem of evaluating the observations using the techniques of statistical testing of the residuals in a mixed linear model is analyzed. In this case, these techniques of testing present particular difficulties, as the quantities that are tested are the total residuals and the separation of the random effects from the errors of the observations cannot be realized. We are also dealing with the problem of the occurrence of unknown variance components, a topic associated usually with the mixed models. Then, and when the degrees of freedom are low, the problem of calculating the degrees of freedom of the distributions used in the above statistical tests arises.



**Keywords:** Mixed linear models, hypothesis testing, degrees of freedom, outliers, residual analysis.

*Whoever knows the ways of Nature will more easily notice her deviations; and, on the other hand, whoever knows her deviations will more accurately describe her way.*

*Francis Bacon, 1620*

## 1. Introduction

In statistics and data analysis, errors, outliers and residuals are three closely related and easily confused measures of the deviation of an observed value, of an observation or an element of statistical sample, from its "theoretical value". The error of an observed value is the deviation of the observed value from the unobservable true value of the observed quantity. Modern developments in the treatment of errors include robust estimation methods and outlier detection methods. An outlier, according to the definition of Hawkins (1980), is an observation, which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism. The diagnostic tool for the detection of outliers is the procedure called analysis of the residuals. The residual of an observed value is the difference between the observed value and the estimation of the true value of the observed quantity. Unfortunately, the residuals are not only related to the existence of errors in the observations, but also to the selection of a proper mathematical model to describe the observations and generally, they are related to all hypotheses that we are making during the analysis of the specific measurements.

Undoubtedly, the outliers have a long history. Obviously, in the words of Francis Bacon, if anyone could identify the "problematic" observations in a data set, then he could understand better the phenomenon that he studies. However, if one was well aware of the phenomenon that he studies, he could easily identify observations that are not in agreement with the other and thus he can improve the estimates and draw more correct conclusions.

The first reference to the outliers seems to be a work of Daniel Bernoulli in 1777, which refers to results obtained by combining more measurements than needed. In his work Bernoulli disagrees to the rejection of a number of measurements on the ground that they initially seem that they don’t agree with the other, a tactic that the astronomers and geodesists of that era were following, such as Roger Joseph Boscovich, who rejected two of the ten measurements of the polar star for the calculation of the Earth's ellipsoid in 1750. It is however a curious feature in geodesy of the late 18th century, while precision in the measurements had become an obsession, no one dealt with the mathematical part of the analysis of measurement errors. The older geodesists did not interpret, neither systematic or random errors, nor the almost inevitable changes in the behavior of the instruments over the years, although it was believed that having many repeated measurements and removing outliers, these errors would be limited. It was destined the great mathematical innovators Andrien Marie Legendre and Carl Friedrich Gauss to develop the method of least squares in the early 19th century and begin the story of outliers' detection

Then, during the 19th century, the first efforts for the development of criteria for the recognition of problematic observations, based on the criterion of least squares, were presented in a work of the American geodesist Benjamin Peirce (1852). It was 75 years after Bernoulli and 45 years after the issue of the method of least squares and the practice of rejecting outliers continued to be common and popular in the analysis of the observations. But the work of Peirce caused a plethora of publications, which were critical of Peirce's criterion for the detection of outliers, but also new criteria appeared. We mention as most important work of Gould (1855), Airy (1856), Wintlock (1856), Chauvenet (1866), Stone (1868), Glaisher (1873, 1874), Edgeworth (1887) and Saunders (1903).

From these articles we distinguish the one of the English geodesist and astronomer George Biddell Airy (1856), who rejected Peirce's criterion, motivated by a scale error in the dimensions of Britain reference ellipsoid, probably caused by retention only of "observations that are in greater agreement between them". Airy argued that, although some observations seem to be outliers, all must be considered and none observation should not be refused if a convincing reason to be able to support the argument that some unusual causes of errors have influenced it, does not exist. Like Bernoulli, he believed that the average of all observations should be taken in order to be exploited all of them with the same weight, as all of them contain valid information. Realizing that the philosophy of doesn't be rejected an observation, matter how big is his difference from the average, in contrast to common sense analysis of the overall behavior of the errors, he didn't reject the observation, but the assumption of normal distribution of errors. Some years before Airy, the German geodesists Bessel and Baeuer (1838) also reported that all observations must be taken into account with the same weight in the calculations and they didn't reject an outlier in the work of connecting the Prussian national network with the Russian one.

Despite efforts to find a criterion for detect and rejection of suspect observations, it had to come in 1960 to be searched the solution in the implementation of the checking the general hypothesis. The precursor of this important chapter of modern statistical analysis, was an article of Karl Pearson (1900). By comparing the observed values by those provided by incidence rates for, Pearson calculated a statistical criterion that follows the same distribution regardless of the type of used data. He could summarize the probability distributions of this statistical criterion at one distribution and use the same tables for each control. This distribution belonged to a group of oblique distributions that had already been defined in 1876 by the German geodesist Friedrich Robert Helmert (1876) using the Greek letter "$\chi$", as Chi Square Distribution. The test has a single parameter, which called degrees of freedom by Ronald Fisher in a paper of 1922, a term used ever since in data analysis to define the available redundant information.

In 1908 William Sealy Gosset introduced the t-distribution for small samples and published his work under the pseudonym "student" while respecting the great teachers of statistic and the criterion t is called since then "student criterion". The term “studentized residuals” was given in honor of this great and modest statistician, because the idea of the division of the residual to its estimated standard deviation is central in the logic of his test.

The most significant work of Ronald Aylmer Fisher (1918) is considered as a milestone in the history of statistics and outliers. Fisher introduced the technique of analysis of variance for the test and separation of "significant" results from random errors. The terms variance and analysis of variance were used for the first time in this work. He also raised the matter of the estimation of the components of the reference variance, which the Fisher named components of variation. In a next article (1924), he presented the chi-square Pearson’s test and t-test of William Gosset in the same frame with the normal distribution and the analysis of variance, using the more general distribution, used by then as F-distribution.

Subsequently, Pole Jerzy Neyman developed the idea of the estimation of the confidence interval and its application on the hypothesis testing and formulated the test of the null hypothesis, in collaboration with Karl Pearson's son, Egon. In developing their theory, Neyman and Egon Pearson recognized the need for the formulation of an alternative hypothesis, and they defined the possible erroneous decisions in checking the null hypothesis. They called first type error the error of rejection of a true hypothesis and second type error the error of the acceptance of a false hypothesis. They paid attention to the possibility of rejection of a hypothesis when it is false. They called this possibility power of test and proposed the term critical region to denote a set of sample statistical values that lead to rejection of the hypothesis being tested. The "area" of the critical region, which they called level of significance, is the probability for the first-type error. On the problem of rejecting outliers perhaps the most important work until that age is the article of Egon Pearson and Chandra Sekar (1936)[1].

Hypothesis testing and analysis of residuals has been introduced to geodesy by the pioneering work of W. Baarda (1968) for the case of the linear model with only deterministic parameters and was later extended by Pope (1976) for the case that the accuracy of the observations is a-priori unknown. During the 80s and 90s, integrated geodesy involves also stochastic parameters and therefore relates to the

[1]. Harter (1978) gives an extensive history of the development of the method of least squares and statistical tests, which refers widely to the issue of outliers’ detection and rejection. Analysis and discussion of methods for identifying and rejecting outliers until recent times is made by Beckman and Cook (1983), while a full report in the literature until 1933 has be made by Rider (1933).

mixed linear model or least squares collocation in geodetic terminology. The problem of hypothesis testing in relation to the integrated geodesy has been studied by Dermanis and Rossikopoulos (1997), in relation to the least squares collocation by Wei (1987), Krakiwsky and Biacs (1990) and in relation to the mixed or random effects linear model by Schaffrin (1987, 1988) and Bock and Schaffrin (1988). The outlier detection procedures can be applied to the error in variable models and total least squares problems seen as mixed linear models, as proposed by Amiri-Simkooei and Jazaeri (2013).

In recent years due to the development of computers and modern forms of geodetic observations, the study of the residuals and outliers has received particular significance under the analysis of data. This is so because a wrong observation can affect significantly the analysis of the observations and may lead to inaccurate results. However, in some cases the outliers are of particular interest as they may not be due to wrong observations and their analysis can lead to new knowledge or discoveries. In this paper the problem of implementation of the test of the general hypothesis to the residuals of the observation is analyzed, in order to identify systematic effects and blunders in analysis with linear mixed model.

## 2. The mixed linear model

Saying linear mixed model in statistics we mean a linear system of equations containing as unknowns fixed effects (or deterministic parameters) and random effects (or stochastic, or random parameters). The random parameters are considered to be normally distributed with zero expectation. These models constitute a generalization of the variance analysis, the principal components analysis and the linear regression methods. They are useful in a wide range of scientific disciplines, in all sciences that are based on observation and measurement or that require methods for evaluation of the measurement uncertainties, such as physics, biology, pharmaceutical, economic and medical sciences in general, in the calibration of instruments and measuring systems, in creating intervals of tolerance in industrial applications and general on laboratory comparisons in metrology, in analyzing spatial data, classified and categorized data, as well as data from temporal and repeated measurements, as for example in geodetic science.

The mixed linear model has the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\,\mathbf{u} + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{y}$ is the $n \times 1$ vector of the observations, $\boldsymbol{\beta}$ the $r \times 1$ vector of fixed effects, $\mathbf{u}$ the $q \times 1$ vector of random effects ( $E\{\mathbf{u}\} = \mathbf{0}$ ) which is accompanied by the covariance matrix $\mathbf{K}$ and $\boldsymbol{\varepsilon}$ the $n \times 1$ vector of random errors ( $E\{\boldsymbol{\varepsilon}\} = \mathbf{0}$ ) which is accompanied by the covariance coefficient matrix $\mathbf{Q}$. The $n \times r$ matrix $\mathbf{X}$ and the

$n \times q$ matrix $\mathbf{Z}$ are the known matrices of coefficients of unknown parameters.

If we call $\mathbf{v} = \mathbf{Z}\,\mathbf{u} + \boldsymbol{\varepsilon}$ the stochastic part of the observation equations, or else the $n \times 1$ vector of marginal residuals, the matrix of their covariances is $\mathbf{M} = \mathbf{Z}\mathbf{K}\,\mathbf{Z}^T + \mathbf{Q}$ and the system of the above equation is written

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} \tag{2}$$

The Best Linear Unbiased Estimation BLUE of the deterministic parameters $\boldsymbol{\beta}$ is given by the relationship

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{M}^{-1}\mathbf{y} \tag{3}$$

which, if we assume that the stochastic variables $u_i$ and the errors $\varepsilon_i$ follow the normal distribution, is the same as the minimum square criterion

$$\mathbf{v}^T\mathbf{M}^{-1}\mathbf{v} = \mathbf{u}^T\mathbf{K}^{-1}\mathbf{u} + \boldsymbol{\varepsilon}^T\mathbf{Q}^{-1}\boldsymbol{\varepsilon} = \min. \tag{4}$$

The Best Linear Unbiased Prediction BLUP of the random (or stochastic) parameters $\mathbf{u}$ follows that it is

$$\begin{aligned}\hat{\mathbf{u}} &= \mathbf{K}\mathbf{Z}^T\mathbf{M}^{-1}\hat{\mathbf{v}} = \mathbf{K}\mathbf{Z}^T(\mathbf{Z}\mathbf{K}\mathbf{Z}^T + \mathbf{Q})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= (\mathbf{Z}^T\mathbf{Q}^{-1}\mathbf{Z} + \mathbf{K}^{-1})^{-1}\mathbf{Z}^T\mathbf{Q}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\end{aligned} \tag{5}$$

as easily demonstrated that $\mathbf{K}\mathbf{Z}^T(\mathbf{Z}\mathbf{K}\mathbf{Z}^T + \mathbf{Q})^{-1} = (\mathbf{Z}^T\mathbf{Q}^{-1}\mathbf{Z} + \mathbf{K}^{-1})^{-1}\mathbf{Z}^T\mathbf{Q}^{-1}$, while the best linear unbiased prediction of the observational errors of the is

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Q}\;\mathbf{M}^{-1}\hat{\mathbf{v}} = \mathbf{Q}\;(\mathbf{Z}\mathbf{K}\mathbf{Z}^T + \mathbf{Q})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\;\hat{\mathbf{u}} \tag{6}$$

From the above relations it also follows that

$$\hat{\mathbf{v}} = (\mathbf{K}\mathbf{Z}^T + \mathbf{Q})\mathbf{M}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{Z}\hat{\mathbf{u}} + \hat{\boldsymbol{\varepsilon}} \tag{7}$$

The covariance coefficients matrices of the estimated parameters $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{u}}$ and $\hat{\boldsymbol{\varepsilon}}$, are given by the relations

$$\begin{aligned}\mathbf{Q}_{\hat{\beta}} &= (\mathbf{X}^T\mathbf{M}^{-1}\mathbf{X})^{-1} = [(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}) - (\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{Z})(\mathbf{Z}^T\mathbf{Q}^{-1}\mathbf{Z} + \mathbf{K}^{-1})^{-1}(\mathbf{Z}^T\mathbf{Q}^{-1}\mathbf{X})]^{-1} \\ \mathbf{Q}_{\hat{u}} &= \mathbf{K}\;\mathbf{Z}^T\mathbf{M}^{-1}(\mathbf{M} - \mathbf{X}(\mathbf{X}^T\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{X}^T)\;\mathbf{M}^{-1}\mathbf{Z}\;\mathbf{K} = \\ &= \mathbf{K}\,\mathbf{Z}^T\mathbf{W}_{\varepsilon\varepsilon}\mathbf{Z}\mathbf{K} = \mathbf{K}\;\mathbf{W}_{uu}\;\mathbf{K} = \\ &= (\mathbf{Z}^T\mathbf{Q}^{-1}\mathbf{Z} + \mathbf{K}^{-1})^{-1}[\mathbf{I} + (\mathbf{Z}^T\mathbf{Q}^{-1}\mathbf{X})\mathbf{Q}_{\hat{\beta}}(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{Z})\,(\mathbf{Z}^T\mathbf{Q}^{-1}\mathbf{Z} + \mathbf{K}^{-1})^{-1}] \\ \mathbf{Q}_{\hat{\beta}\hat{u}} &= -\mathbf{Q}_{\hat{\beta}}(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{Z})\,(\mathbf{Z}^T\mathbf{Q}^{-1}\mathbf{Z} + \mathbf{K}^{-1})^{-1}\end{aligned} \tag{8}$$

and

$$\mathbf{Q}_{\hat{\varepsilon}} = \mathbf{Q}\,\mathbf{M}^{-1}(\mathbf{M} - \mathbf{X}\,(\mathbf{X}^T\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{X}^T)\,\mathbf{M}^{-1}\mathbf{Q} = \mathbf{Q}\,\mathbf{W}_{\varepsilon\varepsilon}\,\mathbf{Q}$$

where $\mathbf{W}_{uu} = \mathbf{Z}^T\mathbf{W}_{\varepsilon\varepsilon}\,\mathbf{Z}$ and the matrix $\mathbf{W}_{\varepsilon\varepsilon}$ is defined from the relation $\hat{\mathbf{v}} = \mathbf{M}\,\mathbf{W}_{\varepsilon\varepsilon}\,\mathbf{v}$ and is given for the GLM ( $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v},\ \mathbf{v}^T\mathbf{M}^{-1}\mathbf{v} = \min.$) as

$$\mathbf{W}_{\varepsilon\varepsilon} = \mathbf{M}^{-1} - \mathbf{M}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{M}^{-1} \tag{9}$$

In linear mixed models, apart from the estimation of unknown parameters $\boldsymbol{\beta}$, the components of the reference variances of the covariance matrices $\mathbf{Q}$ and $\mathbf{K}$ also play an important role. By applying methods of estimating the variance components, the contribution of each random effect to the variance of each observation as well as its significance is estimated. This process is particularly important and fundamental in the analysis of mixed models, as one can determine where he should focus his attention in order to correctly describe the behavior of the random parameters and to correctly select the weight matrix of the observation values.

A useful application of the statistical tests is the detection of one or more $y_i$ observations that do not follow the model and whose removal will lead to a more successful adaptation of it to the other parameters. In general, total residuals $v_i$ or errors $\varepsilon_i$ are often used to evaluate the validity of the assumptions about the statistical and mathematical models and can also be used as tools to identify possible outliers or more generally to identify effects that require more attention.

## 3. Residual analysis in mixed linear models

To detect "problematic observations", those that may have to be removed from the calculations because they contain outliers, we separate three types of residuals that can be checked for possible disturbances. These are (Nobre and Singer 2007):

α. the observational errors (conditional residuals): $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{u}}$

β. the random effects: $\mathbf{Z}\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\varepsilon}}$, and

γ. the marginal residuals: $\hat{\mathbf{v}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Z}\hat{\mathbf{u}} + \hat{\boldsymbol{\varepsilon}}$.

Testing for outliers is an application of the test of the general hypothesis to the stochastic parameters of the linear model and results from a comparison of the estimates $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{v}}$, $\hat{\varphi}$ and $f$ of the initial equations (1)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\,\mathbf{u} + \boldsymbol{\varepsilon} \tag{10}$$

which in this case are the constrained equations for the zero hypothesis ($\boldsymbol{\psi} = \mathbf{0}$) that is tested, with the extended equations

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\,\mathbf{u} + \mathbf{U}\boldsymbol{\psi} + \boldsymbol{\varepsilon} \tag{11}$$

where the $k \times 1$ vector $\boldsymbol{\psi}$ contains the errors of observations that were not taken into account in relation (1) and $\mathbf{U}$ is the $n \times k$ known matrix of their coefficients. These equations lead to new estimates $\hat{\boldsymbol{\beta}}_\delta, \hat{\boldsymbol{\psi}}, \hat{\mathbf{v}}_\delta, \hat{\varphi}_\delta$ and $f_\delta$. In the above equations $\varphi$ denotes the least squares criterion and $f$ the corresponding degrees of freedom. Equivalently, the two previous relationships are written

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} \tag{12}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\psi} + \mathbf{v}$$

The first work in which this extension of the linear equations is proposed as a basis for detecting outliers has been given by Srikantan (1961) and Ferguson (1961).

Initially, let us consider for the sake of simplicity a common reference variance $\hat{\sigma}^2$ for all stochastic parameters according to the relation

$$\mathbf{v} \sim (\mathbf{0}, \sigma^2(\mathbf{Q} + \mathbf{ZKZ}^T)) \quad \text{ή} \quad \mathbf{v} \sim (\mathbf{0}, \sigma^2\mathbf{M}) \tag{13}$$

the estimate of which is calculated by the formula

$$\hat{\sigma}^2 = \frac{\hat{\varphi}}{f} = \frac{\hat{\mathbf{v}}^T\mathbf{M}^{-1}\ \hat{\mathbf{v}}}{tr\{\mathbf{W}_{ææ}\,\mathbf{M}\}} = \frac{\hat{\boldsymbol{\varepsilon}}^T\mathbf{Q}^{-1}\ \hat{\boldsymbol{\varepsilon}} + \hat{\mathbf{u}}^T\mathbf{K}^{-1}\ \hat{\mathbf{u}}}{tr\{\mathbf{W}_{ææ}\,\mathbf{M}\}} \tag{14}$$

where $f = tr\{\mathbf{W}_{ææ}\,\mathbf{M}\}$ are the degrees of freedom. The total significance test

$$H_o : \boldsymbol{\psi} = \mathbf{0} \ \sim \ H_a : \boldsymbol{\psi} \neq \mathbf{0} \tag{15}$$

of the new parameters of the extended linear equations, is an application of the test of the general hypothesis and derived from the comparison of the results of the extended equations (11) with the results of the initial equations (10). This comparison can be derived from the results of the solution only of the extended linear equations

$$F = \frac{1}{k}\hat{\boldsymbol{\psi}}^T\hat{\boldsymbol{\Sigma}}_{\hat{\psi}}^{-1}\hat{\boldsymbol{\psi}} = \frac{\hat{\boldsymbol{\psi}}^T\mathbf{Q}_{\hat{\psi}}^{-1}\hat{\boldsymbol{\psi}}}{k\ \hat{\sigma}_\delta^2} \sim F_{k,f-k} \tag{16}$$

where it was considered that $f_\delta = f - k$, or may result from the comparison of the two solutions

$$F = \frac{f_\delta}{k}\frac{\hat{\varphi} - \hat{\varphi}_\delta}{\hat{\varphi}_\delta} = \frac{f\hat{\sigma}^2 - f_\delta\,\hat{\sigma}_\delta^2}{k\,\hat{\sigma}_\delta^2} \sim F_{k,f-k} \tag{17}$$

or it may result from the solution only of the initial linear equations

$$F = \frac{f-k}{k}\frac{\delta\hat{\varphi}}{\hat{\varphi} - \delta\hat{\varphi}} \sim F_{k,f-k} \tag{18}$$

where applies in any case $\delta\hat{\varphi} = \hat{\boldsymbol{\psi}}^T\mathbf{Q}_{\hat{\psi}}^{-1}\hat{\boldsymbol{\psi}}$, while the estimates $\hat{\boldsymbol{\psi}}$ of the new parameters and their covariance coefficients matrix $\mathbf{Q}_{\hat{\psi}}$ result from the solution of the initial equations $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v}$ according to the equations

$$\begin{aligned}\mathbf{Q}_{\hat{\psi}} &= (\mathbf{U}^T\mathbf{M}^{-1}\mathbf{Q}_{\hat{v}}\ \mathbf{M}^{-1}\mathbf{U})^{-1} = \\ &= [\mathbf{U}^T\mathbf{M}^{-1}(\mathbf{M} - \mathbf{X}(\mathbf{X}^T\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{X}^T)\ \mathbf{M}^{-1}\mathbf{U}]^{-1} = (\mathbf{U}^T\mathbf{W}_{\infty}\mathbf{U})^{-1}\end{aligned} \tag{19}$$

$$\hat{\boldsymbol{\psi}} = \mathbf{Q}_{\hat{\psi}}\ \mathbf{U}^T\mathbf{M}^{-1}\ \hat{\mathbf{v}} \tag{20}$$

Finally, the difference in optimization criterion $\delta\hat{\varphi} = \hat{\varphi} - \hat{\varphi}_{\delta}$ is calculated as

$$\delta\hat{\varphi} = \hat{\mathbf{v}}^T\mathbf{M}^{-1}\mathbf{U}\ (\mathbf{U}^T\mathbf{W}_{\infty}\mathbf{U})^{-1}\mathbf{U}^T\mathbf{M}^{-1}\ \hat{\mathbf{v}} \tag{21}$$

from the results of the solution of the initial equations.

After arithmetic operations, where we considered that

$$\tilde{\sigma}^2 = \hat{\varphi}/(f-k) = f\ \hat{\sigma}^2/(f-k) \tag{22}$$

results that the test for outliers is based on the test statistic

$$T^2 = \frac{\delta\hat{\varphi}}{\tilde{\sigma}^2 k} = \frac{f-k}{f}\frac{\delta\hat{\varphi}}{\hat{\sigma}^2 k} \sim \frac{f\ F_{k,f-k}}{f-k+k\ F_{k,f-k}} \equiv T^2_{k,f-k} \tag{23}$$

calculated from the initial linear equations (1), where the $T^2_{k,f-k}$ distribution could be called a “generalized” Hotelling’s T-square distribution with degrees of freedom $k$ and $f-k$. Alternatively, it is based on the test statistic

$$F = T^2\ \frac{f-k}{f-k\ T^2} \sim F_{k,f-k} \tag{24}$$

The above alternative formulas are based on the relations

$$T^{2(\alpha)}_{k,f-k} = \frac{f\ F^{\alpha}_{k,f-k}}{f-k+k\ F^{\alpha}_{n_2,f-n_2}} \quad \text{or} \quad T^{2(\alpha)}_{k,\nu} = \frac{(\nu+k)\ F^{\alpha}_{k,\nu}}{\nu+k\ F^{\alpha}_{k,\nu}} \tag{25}$$

which links the "generalized" Hotelling distribution with the F distribution.

The degrees of freedom $\nu = f-k$ of the distributions of the above statistical tests are valid if they are small and at the same time if the variance components are neglected considering a common reference variance for all stochastic parameters, according to the relation (13). If we consider that the absolute accuracy of the observations is known, or equivalently, if the degrees of freedom $\nu = f-k$ are very high, (e.g. $\nu > 100$), the test takes the form

$$F = \frac{\delta\hat{\varphi}}{k} \sim F_{k,\infty} \quad \text{or} \quad \chi = \delta\hat{\varphi} \sim \chi^2_k \tag{26}$$

irrespective of the way of calculation of the quantity $\delta\hat{\varphi}$.

The test becomes more compound in the case of unknown absolute accuracy of the stochastic parameters and the simultaneous acceptance of the model of the variance components. In this case, the degrees of freedom $\nu$ depend on the structure of the stochastic model of random parameters and on the way that the variance components are calculated. Their calculation is usually approximated by numerical analysis methods. In this case, a method of calculating the degrees of freedom for the test of general hypothesis is given by Satterthwaite (1946), Giesbrecht and Burns (1985) and Fai and Cornelius (1996). First, let's see the application of the above to the test of a single observation, assuming that all other observations are correct, without errors. The technique of applying the test for outlier detection to a single observation is known in the geodetic literature as "data snooping" (Baarda, 1968)[2].

In the case of testing a single observation, e.g. of $y_i$, the extended equations are of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}_i\psi_i + \boldsymbol{\varepsilon} \tag{27}$$

or equivalently, the initial and the extended system of linear equations are written

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{v} \\ \mathbf{y} &= \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}_i\,\psi + \mathbf{v} \end{aligned} \tag{28}$$

where $\psi_i$ is the error of the $y_i$ observation, which was ignored in previous calculations of the initial equations and $\mathbf{e}_i$ is the $i$ column of the matrix $\mathbf{I}_n$. According to the above, this comparison can be derived from the results of the solution of the initial system of linear equations (2) or (9) only, or it can result from solving the extended equations (27) only, or by comparing the results of the solutions of the two systems. We will deal with the first case only, the application of statistical tests to the results of the solution of the initial equations.

In this case and when the marginal residuals $\mathbf{v}$ of the initial equations $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v}$ are testing, the modified residuals are calculated (Dermanis, 1987)

$$\hat{\boldsymbol{\xi}} = \mathbf{M}^{-1}\hat{\mathbf{v}} \tag{29}$$

where $\hat{\mathbf{v}} = \mathbf{Z}\hat{\mathbf{u}} + \hat{\boldsymbol{\varepsilon}} = \mathbf{M}\mathbf{W}_{\varepsilon\varepsilon}\,\mathbf{v}$ and $\mathbf{W}_{\varepsilon\varepsilon} = \mathbf{M}^{-1} - \mathbf{M}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{M}^{-1}$.

From the above relation, if the law of variance-covariance propagation is applied, the coefficient matrix of variances and covariances of quantities $\hat{\boldsymbol{\xi}}$ results

---

2. The term data snooping appears in the statistical literature before Baarda (1968). It is an individual statistical process in the more general procedure of data analysis called Data-Dredging: Snooping is the process of testing from the data all of a predesignated (though possibly infinite) set of hypotheses (Selvin ans Stuart, 1966).

$$\mathbf{Q}_{\hat{\xi}} = \mathbf{M}^{-1}\mathbf{Q}_{\hat{v}}\ \mathbf{M}^{-1} = \mathbf{M}^{-1}(\mathbf{M} - \mathbf{X}(\mathbf{X}^T\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{X}^T)\ \mathbf{M}^{-1} = \mathbf{W}_{\hat{e}\hat{e}} \tag{30}$$

where $\mathbf{Q}_{\hat{v}} = \mathbf{M} - \mathbf{X}(\mathbf{X}^T\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{X}^T$. According to the equations (19) and (20), it follows that

$$\hat{\psi}_i = \frac{\hat{\xi}_i}{[\mathbf{Q}_{\hat{\xi}}]_{ii}} \quad \text{and} \quad Q_{\hat{\psi}}^{-1} = 1/q^2(\hat{\psi}_i) = [\mathbf{Q}_{\hat{\xi}}]_{ii} = [\mathbf{W}_{\hat{e}\hat{e}}]_{ii} \tag{31}$$

The test of the $y_i$ observation is therefore based on the studentized error

$$\tau_i = \frac{\hat{\xi}_i}{\hat{\sigma}(\hat{\xi}_i)} \sim \tau_\nu \quad , \quad \hat{\sigma}(\hat{\xi}_i) = \sqrt{(\hat{\mathbf{W}}_{\hat{e}\hat{e}})_{ii}} \tag{32}$$

which follows the $\tau_\nu$ (tau) distribution with $\nu$ degrees of freedom. The tau distribution was devised by W. R. Thompson (1935) and became known in the geodetic literature by A. J. Pope (1976), who formulated the above-mentioned form of statistical test.

Regarding the degrees of freedom $\nu$ of the above distribution, they are calculated as the case may be, according to the following:

a. For very high degrees of freedom (e.g. $f > 100$) or for known accuracy of observations, it can be assumed that $\nu \to \infty$ and the tau distribution is identical to the standard normal distribution $N(0,1)$. The test has been given in this form in the pioneering work of Baarda (1968) with title “a testing procedure in use in geodetic networks”, who used the term "data snooping" for the technique with which the test is applied.
b. For low degrees of freedom $f$ and for a common variance of unit weight $\sigma^2$ according to the relationship (13), they are given as $\nu = f$ (Pope 1976, Kok 1984).
c. For low degrees of freedom $f$ and the more complex case of the unknown variance components, the degrees of freedom $\nu = \nu_i$ results, in a very good approximation, that are (Satterthwaite, 1946, Giesbrecht and Burns, 1985)

$$\nu_i \approx \frac{2\hat{\sigma}_{\xi}^4}{Var(\hat{\sigma}_{\xi}^2)} = \frac{2\hat{\sigma}_{\xi}^4}{\hat{\sigma}^2(\hat{\sigma}_{\xi}^2)} \tag{33}$$

where $\hat{\sigma}_{\xi}^2 = \hat{\sigma}^2(\hat{\xi}_i)$ is the variance of the modified residual $\hat{\xi}_i$ and $\hat{\sigma}^2(\hat{\sigma}_{\xi}^2) = Var(\hat{\sigma}_{\xi}^2)$ the variance of the variance $\hat{\sigma}_{\xi}^2$, is being calculated by applying the low of variance propagation for the estimation of the relationships of variance components.[3]

---

[3]. For example, in testing the difference between the mean of two samples, of different and un-

The values of tau distribution relate to the values of t distribution through the relation

$$\tau_{\nu}^{\alpha/2} = \sqrt{\frac{\nu\ (t_{\nu-1}^{\alpha/2})^2}{\nu - 1 + (t_{\nu-1}^{\alpha/2})^2}} = \sqrt{\frac{\nu\ F_{1,\nu-1}^{\alpha}}{\nu - 1 + F_{1,\nu-1}^{\alpha}}} \tag{34}$$

Reversing this relation, results in

$$t_{\nu-1}^{\alpha/2} = \tau_{\nu}^{\alpha/2} \sqrt{\frac{\nu - 1}{\nu - (\tau_{\nu}^{\alpha/2})^2}} \tag{35}$$

from which the alternative form of control arises

$$\tau_i = \frac{\hat{\xi}_i}{\hat{\sigma}(\hat{\xi}_i)}\ , \quad t_i = \tau_i \sqrt{\frac{\nu - 1}{\nu - \tau_i^2}} \sim t_{\nu-1} \tag{36}$$

This statistical test has been given in this form by Beckman and Trussell (1974), Weisberg (1980) and in Greek literature by Dermanis (1987). The quantity $\tau_i$ is called internal studentized error, while the $t_i$ external studentized error, because it can also arise from the relationship

$$t_i = \frac{\tilde{\tilde{\xi}}_i}{\tilde{\sigma}_{(i)}(\tilde{\tilde{\xi}}_i)} = \frac{\tilde{\tilde{v}}_i}{\hat{\sigma}_{(-i)}\ q_{ii}} \sim t_{\nu-1} \tag{37}$$

where the modified residual $\tilde{\tilde{\xi}}_i$ and its variance $\tilde{\sigma}_{(i\}}^2(\tilde{\tilde{\xi}}_i)$ results as a prediction from the solution of the initial linear mixed model $\mathbf{y}_{(-i)} = \mathbf{X}_{(-i)}\boldsymbol{\beta}_{(-i)} + \mathbf{v}_{(-i)}$, i.e. without the contribution of observation $y_i$ . If we write the matrix $\mathbf{M}$ in the form

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{(-i)} & \mathbf{m}_{(-i)} \\ \mathbf{m}_{(-i)}^T & m_{ii} \end{bmatrix} \quad \text{και} \quad \begin{bmatrix} \tilde{\boldsymbol{\xi}}_{(-i)} \\ \tilde{\tilde{\xi}}_i \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{(-i)} & \mathbf{m}_{(-i)} \\ \mathbf{m}_{(-i)}^T & m_{ii} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{v}}_{(-i)} \\ \tilde{v} \end{bmatrix} \tag{38}$$

where $\tilde{v}_i = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{(-i)}$, then the relations apply

$$\tilde{\tilde{\xi}}_i = \tilde{m}_{ii}^{-1}\, \tilde{\tilde{v}}_i\ , \quad \tilde{\sigma}_{(-i)}^2(\tilde{\tilde{\xi}}_i) = \tilde{m}_{ii}^{-2}\, \tilde{\sigma}_{(-i)}^2(\tilde{\tilde{v}}_i) = \tilde{m}_{ii}^{-2}\, \hat{\sigma}_{(-i)}^2\, q_{ii}^2$$

$$\tilde{\tilde{v}}_i = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{(-i)} - \mathbf{m}_{(-i)}^T \mathbf{M}_{(-i)}^{-1} \hat{\mathbf{v}}_{(-i)} = \tilde{v}_i - \mathbf{m}_{(-i)}^T \mathbf{M}_{(-i)}^{-1} \hat{\mathbf{v}}_{(-i)}$$

$$\tilde{\sigma}_{(-i)}^2(\tilde{\tilde{v}}_i) = \hat{\sigma}_{(-i)}^2\, q_{ii}^2 = \hat{\sigma}_{(-i)}^2\, \tilde{m}_{ii}^{-2}\, (m_{ii} - \mathbf{g}^T\, (\mathbf{X}_{(-i)}^T\ \mathbf{M}_{(-i)}^{-1}\ \mathbf{X}_{(-i)})^{-1} \mathbf{g}) \tag{39}$$

---

known accuracy of their values, the degrees of freedom of t-distribution become according to the formula of Satterthwaite $\nu = (s_1^2 / n_1 + s_2^2 / n_2) / [s_1^4 / n_1^2 (n_1 - 1) + s_2^4 / n_2^2 (n_2 - 1)]$.

and

$$\tilde{m}_{ii} = m_{ii} - \mathbf{m}_{(-i)}^T \mathbf{M}_{(-i)}^{-1} \mathbf{m}_{(-i)}, \quad \mathbf{g} = \mathbf{x}_i - \mathbf{m}_{(-i)}^T \mathbf{M}_{(-i)}^{-1} \mathbf{X}_{(-i)} \tag{40}$$

For uncorrelated observations the external studentized error becomes[4]

$$t_i = \frac{\tilde{v}_i}{\hat{\sigma}_{(-i)}\, q_{ii}} \sim t_{\nu-1} \tag{41}$$

where $\tilde{v}_i = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{(-i)}$ is the prediction of error, and

$$\tilde{\sigma}_{(-i)}^2(\tilde{v}_i) = \hat{\sigma}_{(-i)}^2\, q_{ii}^2 = \hat{\sigma}_{(-i)}^2\, (m_{ii}^{-1} - m_{ii}^{-2}\, \mathbf{x}_i^T\, (\mathbf{X}_{(-i)}^T\, \mathbf{M}_{(-i)}^{-1}\, \mathbf{X}_{(-i)})^{-1} \mathbf{x}_i) \tag{42}$$

its variance.

Where a set of observations is tested (the set with index 2, $\mathbf{y}_2 = \mathbf{X}_2\, \boldsymbol{\beta} + \mathbf{v}_2$), affected each of them by a simple error $\psi_i$, with respect to the other observations that are considered to be unmistakable, without errors (to set index 1, $\mathbf{y}_1 = \mathbf{X}_1\, \boldsymbol{\beta} + \mathbf{v}_1$), the test can be done either by the results of the simultaneous analysis of the observations of the two sets, or by the results of the analysis of the first set of observations and applying prediction methods for calculating the errors and their covariance matrix of the second set. The two methods are equivalent to each other, as in the case of one suspicious observation discussed above. The extended system of linear equations is written

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} \boldsymbol{\psi}_2 + \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} \tag{43}$$

where the vector $\boldsymbol{\psi}_2$ contains the errors $\psi_i$ of observations of the set (2), and corresponding to the above the following quantities are calculated

$$\mathbf{Q}_{\hat{\psi}} = [\mathbf{W}_{\alpha\epsilon}^{-1}]_{22} = \mathbf{W}_{22}^{-1} \tag{44}$$

where the matrix $\mathbf{W}_{22}^{-1} = [\mathbf{W}_{\alpha\epsilon}^{-1}]_{22}$ is a submatrix of the inverse $\mathbf{W}_{\alpha\epsilon}^{-1}$ which corresponds to the observations of the set (2)

$$\hat{\boldsymbol{\psi}}_2 = \mathbf{Q}_{\hat{\psi}}\, [\mathbf{M}^{-1}\hat{\mathbf{v}}]_2 \tag{45}$$

and

$$\delta\hat{\varphi} = \hat{\boldsymbol{\psi}}^T \mathbf{Q}_{\hat{\psi}}^{-1}\, \hat{\boldsymbol{\psi}} == \hat{\boldsymbol{\xi}}_2^T\, \mathbf{Q}_{\hat{\psi}}\, \hat{\boldsymbol{\xi}}_2 = \hat{\boldsymbol{\xi}}_2^T [\mathbf{W}_{\alpha\epsilon}^{-1}]_{22} \hat{\boldsymbol{\xi}}_2 \tag{46}$$

4. As it was proven by Beckman and Trussell (1974) and Cook and Weisberg (1982), for the case of uncorrelated observations.

where $\hat{\boldsymbol{\xi}}_2 = [\mathbf{M}^{-1}\,\hat{\mathbf{v}}]_2$ are the modified total residuals of the observations being tested, or alternatively if the operations will be done

$$\hat{\boldsymbol{\xi}}_2 = (\mathbf{M}_{22} - \mathbf{M}_{12}^T\,\mathbf{M}_{11}^{-1}\,\mathbf{M}_{12})^{-1}(\hat{\mathbf{v}}_2 - \mathbf{M}_{12}^T\,\mathbf{M}_{11}^{-1}\,\hat{\mathbf{v}}_1) \tag{47}$$

where $\mathbf{M}_{11}$, $\mathbf{M}_{22}$, $\mathbf{M}_{12}$ are submatrices of $\mathbf{M}$.

In the first case, based on the simultaneous analysis of the observations of both sets, the modified total residuals $\hat{\boldsymbol{\xi}}_2$ are calculated and the test is based on the relations:

a. For a very large number of degrees of freedom (e.g. $f > 100$) or, for known accuracy

$$\chi^2 = \hat{\boldsymbol{\xi}}_2^T\;\mathbf{W}_{22}^{-1}\;\hat{\boldsymbol{\xi}}_2 \sim \chi^2_{n_2} \tag{48}$$

b. For low degrees of freedom and an unknown reference variance, common for all stochastic parameters

$$T^2 = \frac{\hat{\boldsymbol{\xi}}_2^T\,\mathbf{W}_{22}^{-1}\,\hat{\boldsymbol{\xi}}_2}{n_2\;\hat{\sigma}^2} \sim T^2_{n_2,\,f-n_2} \tag{49}$$

where $T^2_{n_2,\,f-n_2}$ is the “generalized” Hotelling distribution, which is a generalization of the $t$ distribution in the analysis of many variables, or alternatively

$$T^2 = \frac{\hat{\boldsymbol{\xi}}_2^T\,\mathbf{W}_{22}^{-1}\;\hat{\boldsymbol{\xi}}_2}{n_2\;\hat{\sigma}^2}\;\;,\quad F = T^2\,\frac{f - n_2}{f - n_2\;T^2} \sim F_{n_2,\,f-n_2} \tag{50}$$

c. For low degrees of freedom and the more complicated case of the unknown variance components

$$T^2 = \frac{\hat{\boldsymbol{\xi}}_2^T\,\hat{\mathbf{W}}_{22}^{-1}\;\hat{\boldsymbol{\xi}}_2}{n_2} \sim T_{n_2,\nu} \tag{51}$$

where the matrix $\hat{\mathbf{W}}_{\varepsilon\varepsilon}$ results from the iterations in the course of estimating the unknown variance components, or alternatively

$$T^2 = \frac{\hat{\boldsymbol{\xi}}_2^T\,\hat{\mathbf{W}}_{22}^{-1}\;\hat{\boldsymbol{\xi}}_2}{n_2}\;\;,\quad F = T^2\,\frac{\nu}{(\nu + n_2) - n_2\;T^2} \sim F_{n_2,\nu} \tag{52}$$

The above $\nu$ degrees of freedom of the $F_{n_2,\,\nu}$ distribution are calculated by the equations (Fai and Cornelius, 1996)

$$v = \frac{2E}{E-k} \;\;,\;\; E = \sum_{i=1}^{k} \frac{v_i}{v_i - 2} \ell_{[v_i > 2]} \tag{53}$$

where $v_i$ denotes the degrees of freedom for each $i$ observation from the set of $n_2$ being tested, calculated according to formula (33) and the coefficient $\ell_{[v_i>2]}$ indicates that only the terms $v_i > 2$ are summed.

For the second case, assuming that we want to test the second set of observations from the solution of the first set, the test is based on the alternative form of the statistic

$$F = \frac{\tilde{\boldsymbol{\xi}}_2^T \mathbf{Q}_{\tilde{\xi}}^{-1} \tilde{\boldsymbol{\xi}}_2}{n_2 \; \hat{\sigma}_1^2} \sim F_{n_2, f-n_2} \tag{54}$$

where $\tilde{\boldsymbol{\xi}}_2 = \tilde{\boldsymbol{\xi}}_2^{(1)}$ is the prediction of the modified residuals of the second set of observations

$$\tilde{\boldsymbol{\xi}}_2^{(1)} = (\mathbf{M}_{22} - \mathbf{M}_{12}^T \mathbf{M}_{11}^{-1} \mathbf{M}_{12})^{-1} (\tilde{\mathbf{v}}_2^{(1)} - \mathbf{M}_{12}^T \mathbf{M}_{11}^{-1} \hat{\mathbf{v}}_1) = \tilde{\mathbf{M}}_{22}^{-1} (\tilde{\mathbf{v}}_2^{(1)} - \mathbf{M}_{12}^T \mathbf{M}_{11}^{-1} \hat{\mathbf{v}}_1) \tag{55}$$

where $\tilde{\mathbf{M}}_{22} = \mathbf{M}_{22} - \mathbf{M}_{12}^T \mathbf{M}_{11}^{-1} \mathbf{M}_{12}$ and

$$\mathbf{Q}_{\tilde{\xi}} = \tilde{\mathbf{M}}_{22}^{-1} [\mathbf{M}_{22} - \mathbf{G} (\mathbf{X}_1^T \; \mathbf{M}_{11}^{-1} \; \mathbf{X}_1)^{-1} \mathbf{G}^T ] \tilde{\mathbf{M}}_{22}^{-1} \tag{56}$$

their covariance coefficient matrix, and

$$\mathbf{G} = \mathbf{X}_2 - \mathbf{M}_{12}^T \mathbf{M}_{11}^{-1} \mathbf{X}_1 \tag{57}$$

From the analysis of the first set of observations, the estimates $\hat{\boldsymbol{\beta}}^{(1)}$ of the unknown parameters and the coefficient matrix $\mathbf{Q}_{\hat{\beta}}$ of their covariances are given by

$$\hat{\boldsymbol{\beta}}^{(1)} = (\mathbf{X}_1^T \mathbf{M}_{11}^{-1} \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{M}_{11}^{-1} \; \mathbf{y}_1 \tag{58}$$

$$\mathbf{Q}_{\hat{\beta}} = (\mathbf{X}_1^T \; \mathbf{M}_{11}^{-1} \; \mathbf{X}_1)^{-1} \tag{59}$$

as well as the prediction of the total residuals of the second set

$$\tilde{\mathbf{v}}_2^{(1)} = \mathbf{y}_2 - \mathbf{X}_2 \, \hat{\boldsymbol{\beta}}^{(1)} \tag{60}$$

where $\mathbf{M}_{11}$, $\mathbf{M}_{22}$, $\mathbf{M}_{12}$ are the submatrices of $\mathbf{M}$.

It is easy to prove that $\tilde{\boldsymbol{\xi}}_2^T \mathbf{Q}_{\tilde{\xi}}^{-1} \tilde{\boldsymbol{\xi}}_2 = \hat{\tilde{\mathbf{v}}}_2^T \mathbf{Q}_{\hat{\tilde{v}}}^{-1} \hat{\tilde{\mathbf{v}}}_2$, where

$$\hat{\tilde{\mathbf{v}}}_2 = \tilde{\mathbf{v}}_2^{(1)} - \mathbf{M}_{12}^T \mathbf{M}_{11}^{-1} \hat{\mathbf{v}}_1 \tag{61}$$

and

$$\mathbf{Q}_{\tilde{\tilde{v}}} = \mathbf{M}_{22} - \mathbf{G}(\mathbf{X}_1^T \mathbf{M}_{11}^{-1} \mathbf{X}_1)^{-1} \mathbf{G}^T \tag{62}$$

If the two sets of observations (1) and (2) are uncorrelated with each other, then $\mathbf{M}_{12} = \mathbf{0}$, $\tilde{\mathbf{M}}_{22} = \mathbf{M}_{22}$ and

$$\tilde{\tilde{\mathbf{v}}}_2 = \tilde{\mathbf{v}}_2^{(1)} \text{ and } \mathbf{Q}_{\tilde{\tilde{v}}} = \mathbf{M}_{22} - \mathbf{X}_2(\mathbf{X}_1^T \mathbf{M}_{11}^{-1} \mathbf{X}_1)^{-1} \mathbf{X}_2^T \tag{63}$$

It is proved (Dermanis and Rossikopoulos, 1991) that this statistical test is equivalent to the test of so-called stochastic linear hypotheses, an idea introduced by Schaffrin (1987). This is also exactly one of statistics derived by Wei (1987) for the simpler case of “group-wise collocation”

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{X}_1 \boldsymbol{\beta} + \mathbf{u}_1 + \boldsymbol{\varepsilon}_1 \\ \mathbf{y}_2 &= \mathbf{X}_2 \boldsymbol{\beta} + \mathbf{u}_2 + \boldsymbol{\varepsilon}_2 \end{aligned} \tag{64}$$

Large values of the test statistic $F$ or $T^2$ for many of the observations do not necessarily indicate the presence of multiple gross errors, especially for the high quality studies. Very often, they are probably due to modeling errors on the part of the deterministic or stochastic parameters, or in covariance model of stochastic parameters. Tests for multiple outliers on observations can be performed as described above, but their interpretation as gross errors are again not possible for the usual high quality of the observations. Such detected outliers may signify also the existence of model inconsistency between the two sets of observations, those with outliers and those without.

In this case, when the compatibility of two sets of observations is being tested, the problem is how the original data set is going to be divided into two subsets. Therefore, the test should be applied only when a natural separation exists. This is the case with prior information from a previous data analysis which is treated through the device of "pseudo-observations". Another such case appears when observations in a network are to be combined with separate gravity related observations in the surrounding area. An example is the incorporation of independent gravity and height information with observations in a GPS network, which is necessary for the determination of orthometric rather than ellipsoidal heights.

## 4. Conclusions

It could be said that the above technique of testing could be applied exactly in the same manner to the errors $\hat{\boldsymbol{\varepsilon}}$, or to the random effects $\hat{\mathbf{e}} = \mathbf{Z}\hat{\mathbf{u}}$. But in any case, either the total residuals $\hat{\mathbf{v}}$, the errors $\hat{\boldsymbol{\varepsilon}}$, or the stochastic parameters $\hat{\mathbf{u}}$ are tested applying the procedure described above, the significance of the new parameters $\boldsymbol{\psi}$ is being tested in fact, as in each case the results of the extended equations

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\psi} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \tag{65}$$

are compared in the same way with the results of the initial equations

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \tag{66}$$

Let's see the implementation of statistical testing for the detection of outliers in the errors $\hat{\boldsymbol{\varepsilon}}$, in the general case where the observations are correlated with each other. The modified errors $\hat{\boldsymbol{\xi}}$ and their covariance matrix $\mathbf{Q}_{\hat{\xi}}$ are calculated

$$\begin{aligned} \hat{\boldsymbol{\xi}} &= \mathbf{Q}^{-1}\,\hat{\boldsymbol{\varepsilon}} = \mathbf{Q}^{-1}\,\mathbf{Q}\,\mathbf{M}^{-1}\,\hat{\mathbf{v}} = \mathbf{M}^{-1}\,\hat{\mathbf{v}} \\ \mathbf{Q}_{\hat{\xi}} &= \mathbf{Q}^{-1}\,\mathbf{Q}_{\hat{\varepsilon}}\,\mathbf{Q}^{-1} = \mathbf{W}_{\varpi} \end{aligned} \tag{67}$$

where we also took into account that $\hat{\boldsymbol{\varepsilon}} = \mathbf{Q}\,\mathbf{M}^{-1}\,\hat{\mathbf{v}}$. The estimates of the additional parameters and their covariance matrix appear to be

$$\begin{aligned} \hat{\boldsymbol{\psi}} &= \mathbf{Q}_{\hat{\psi}}\,\mathbf{U}^T\mathbf{M}^{-1}\,\hat{\mathbf{v}} = \mathbf{Q}_{\hat{\psi}}\,\mathbf{U}^T\mathbf{Q}^{-1}\,\hat{\boldsymbol{\varepsilon}} = \mathbf{Q}_{\hat{\psi}}\,\mathbf{U}^T\hat{\boldsymbol{\xi}} \\ \mathbf{Q}_{\hat{\psi}} &= (\mathbf{U}^T\mathbf{W}_{\varpi}\mathbf{U})^{-1} \end{aligned} \tag{68}$$

If we take into account that

$$\mathbf{U}\,(\mathbf{U}^T\mathbf{W}_{\varpi}\mathbf{U})^{-1}\,\mathbf{U}^T = \mathbf{W}_{\varpi}^{-1} \quad \text{and} \quad \mathbf{Q}_{\hat{\xi}}^{-1} = \mathbf{W}_{\varpi}^{-1} = \mathbf{Q}\,\mathbf{Q}_{\hat{\varepsilon}}^{-1}\mathbf{Q}$$

then it follows that

$$\delta\hat{\varphi} = \hat{\boldsymbol{\psi}}^T\mathbf{Q}_{\hat{\psi}}^{-1}\,\hat{\boldsymbol{\psi}} = \hat{\boldsymbol{\xi}}^T\mathbf{W}_{\varpi}^{-1}\,\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\varepsilon}}^T\mathbf{Q}^{-1}\mathbf{W}_{\varpi}^{-1}\,\mathbf{Q}^{-1}\,\hat{\boldsymbol{\varepsilon}} = \hat{\boldsymbol{\varepsilon}}^T\mathbf{Q}_{\hat{\varepsilon}}^{-1}\,\hat{\boldsymbol{\varepsilon}} \tag{69}$$

In the same way the test of the random effects $\hat{\mathbf{e}} = \mathbf{Z}\hat{\mathbf{u}}$ is based on their modified values and their covariance matrix

$$\begin{aligned} \hat{\boldsymbol{\xi}} &= \mathbf{Q}_e^{-1}\,\hat{\mathbf{e}} = (\mathbf{Z}\,\mathbf{K}\,\mathbf{Z}^T)^{-1}\mathbf{Z}\,\hat{\mathbf{u}} = (\mathbf{Z}\,\mathbf{K}\,\mathbf{Z}^T)^{-1}\mathbf{Z}\,\mathbf{K}\,\mathbf{Z}^T\mathbf{M}^{-1}\,\hat{\mathbf{v}} = \mathbf{M}^{-1}\,\hat{\mathbf{v}} \\ \mathbf{Q}_{\hat{\xi}} &= (\mathbf{Z}\mathbf{K}\,\mathbf{Z}^T)^{-1}\mathbf{Z}\mathbf{Q}_{\hat{u}}\,\mathbf{Z}^T(\mathbf{Z}\mathbf{K}\,\mathbf{Z}^T)^{-1} = \mathbf{W}_{\varpi} \end{aligned} \tag{70}$$

where we have taken the relationship (7) into account, $\mathbf{Q}_{\hat{u}} = \mathbf{K}\,\mathbf{Z}^T\mathbf{W}_{\varpi}\,\mathbf{Z}\mathbf{K}$. It is easy to be proved that

$$\delta\hat{\varphi} = \hat{\boldsymbol{\psi}}^T\mathbf{Q}_{\hat{\psi}}^{-1}\,\hat{\boldsymbol{\psi}} = \hat{\mathbf{u}}^T\mathbf{Z}^T(\mathbf{Z}\,\mathbf{Q}_{\hat{u}}\,\mathbf{Z}^T)^{-1}\mathbf{Z}\,\hat{\mathbf{u}} = \hat{\boldsymbol{\xi}}^T\,\mathbf{W}_{\varpi}^{-1}\,\hat{\boldsymbol{\xi}} \tag{71}$$

where

$$\hat{\boldsymbol{\psi}} = \mathbf{Q}_{\hat{\psi}} \mathbf{U}^T (\mathbf{Z}\mathbf{K}\mathbf{Z}^T)^{-1} \mathbf{Z}\hat{\mathbf{u}} = \mathbf{Q}_{\hat{\psi}} \mathbf{U}^T \mathbf{M}^{-1} \hat{\mathbf{v}}, \ \mathbf{Q}_{\hat{\psi}} = (\mathbf{U}^T \mathbf{W}_{\varepsilon\varepsilon} \mathbf{U})^{-1} \quad (72)$$

We have taken into account that $(\mathbf{Z}\mathbf{K}\mathbf{Z}^T)^{-1} \mathbf{U}\mathbf{Q}_{\hat{\psi}} \mathbf{U}^T (\mathbf{Z}\mathbf{K}\mathbf{Z}^T)^{-1} =$

$= (\mathbf{Z}\mathbf{K}\mathbf{Z}^T)^{-1} \mathbf{W}_{\varepsilon\varepsilon}^{-1} (\mathbf{Z}\mathbf{K}\mathbf{Z}^T)^{-1} = (\mathbf{Z}\mathbf{K}\mathbf{Z}^T \mathbf{W}_{\varepsilon\varepsilon} \mathbf{Z}\mathbf{K}\mathbf{Z}^T)^{-1} = (\mathbf{Z}\mathbf{Q}_{\hat{u}}\mathbf{Z}^T)^{-1}$

The following proves to be true

$$\delta\hat{\phi} = \hat{\boldsymbol{\psi}}^T \mathbf{Q}_{\hat{\psi}}^{-1} \hat{\boldsymbol{\psi}} = \hat{\boldsymbol{\xi}}^T \mathbf{W}_{\varepsilon\varepsilon}^{-1} \hat{\boldsymbol{\xi}} = \hat{\mathbf{u}}^T \mathbf{Q}_{\hat{u}}^{-1} \hat{\mathbf{u}} \quad (73)$$

where we have accepted that $\mathbf{Z}^T (\mathbf{Z}\mathbf{Q}_{\hat{u}} \mathbf{Z}^T)^{-1} \mathbf{Z} = \mathbf{Q}_{\hat{u}}^{-1}$. We conclude in the same quantity $\delta\hat{\varphi}$ in any case, which turns out to be

$$\delta\hat{\phi} = \hat{\boldsymbol{\psi}}^T \mathbf{Q}_{\hat{\psi}}^{-1} \hat{\boldsymbol{\psi}} = \hat{\boldsymbol{\varepsilon}}^T \mathbf{Q}_{\hat{\varepsilon}}^{-1} \hat{\boldsymbol{\varepsilon}} = \hat{\mathbf{v}}^T \mathbf{Q}_{\hat{v}}^{-1} \hat{\mathbf{v}} = \hat{\mathbf{e}}^T \mathbf{Q}_{\hat{e}}^{-1} \hat{\mathbf{e}} = \hat{\mathbf{u}}^T \mathbf{Q}_{\hat{u}}^{-1} \hat{\mathbf{u}} \quad (74)$$

In these three cases, there are essentially three alternative forms of the same statistical test: the significance test of the additional parameters $\boldsymbol{\psi}$.

Generally, confusion appears on this issue in the international literature. For example, if the test of a set $\mathbf{u}_i$ of stochastic parameters based on quantity $\hat{\mathbf{u}}_i^T [\mathbf{Q}_{\hat{u}}^{-1}]_{ii} \hat{\mathbf{u}}_i$ fails, then this does not necessarily indicate an unsuccessful selection of random effects parameters or an incorrect selection of their covariance matrix, but may be due for example to gross errors that affect an observation or to a wrong selection of the covariance matrix for the errors $\boldsymbol{\varepsilon}$. In the same way, the possible existence of outliers in the implementation of statistical test referred to errors $\boldsymbol{\varepsilon}$ cannot be interpreted as gross errors in observations. However, one could detect outliers in stochastic quantities $\hat{u}_i$ by analyzing graphs of these values or of the quantities $\hat{e}_i$ in terms of the sizes they describe, or by analyzing the elements of their covariance matrix or the graphs of normalized values $\hat{u}_i / \hat{\sigma}(u_i)$, or one could have a picture of the stochastic behavior of stochastic parameters by further analysis of total residuals $\hat{v}_i$.

## References


Amiri-Simkooei, A. R. and S. Jazaeri (2013). Data snooping procedure applied to errors-in-variables models. *Stud. Geophys. Geod.*, 57, 426-441.

Airy, G. B. (1856). Letter from Professor Airy, Astronomer Royal, to the Editor. *Astronomical Journal*, 90, 137-138.

Baarda, W. (1968). *A Testing Procedure in Use in Geodetic Networks*. Netherlands Geodetic Commission, Vol. 2, Number 5.

Beckman, R. J. and R. D. Cook (1983). Outlier … s. *Technometrics*, vol. 25, No. 2, 119 – 163.

Beckman, R. J. and H. J. Trussell (1974). The Distribution of an Arbitrary Studentized Residual and the Effects of Updating in Multiple Regression. *Journal oft he American Statistical Association*, 69, 199-201.

Bessel, F. W and J. J. Baeuer (1838). *Gradmessung in Ostpreussen und ihre Verbindung mit Preussischen und Russischen Dreieckskketten*. Berlin: Dümmler.

Breslow, N. E. and D. G. Clayton (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, March 1993, 9-25.

Chauvenet, W, (1860). *A manual of Spherical and Practical Astronomy* (Vol. II, 5th ed.). New York: Dover.

Cook, R. D. and S. Weisberg (1982). *Residuals and Influence in Regression*. New York: Chapman and Hall.

Dermanis, A. (1987). *Adjustment of observations and estimation theory*. Editions Ziti (in Greek).

Dermanis, A. and D. Rossikopoulos (1997). Statistical Inference in Integrated Geodesy. *IUGG XXth General Assembly*. Vienna, August 11-24, 1997.

Dermanis, A. and R. Rummel (2000). Data analysis methods in geodesy. In: A. Dermanis, A. Grun and F. Sanso (eds.). *Geomatic Methods for the Analysis of Data in the Earth Sciences*. Springer Verlag, 17-92.

Edgeworth, F. Y. (1887). On Discordant Observations. *Philosophical Magazine*, 23, Ser. 5, 364-375.

Elston, D. A. (1998). Estimation of denominator degrees of freedom of F-Distributions for assessing Wald Statistics for fixed-effect in unbalanced mixed models. *Biometrics* 54, 1085-1096.

Fai, A.H.T. and Cornelius, P.L. (1996). Approximate F-tests of Multiple Degree of Freedom Hypotheses in Generalized Least Squares Analyses of Unbalanced Split-plot Experiments. *Journal of Statistical Computing and Simulation* 54, 363-378.

Ferguson, T. S. (1961). On the Rejection of Outliers. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol.1), University of California Press, 253-287.

Fisher, RA (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 52 (2): 399–433.

Fisher, RA (1922). On the Interpretation of $\chi 2$ from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society* 85 (1): 87–94

Fisher, RA (1924). On a Distribution Yielding the Error Functions of Several Well Known Statistics Proceedings of the International Congress of Mathematics, Toronto, 2: 805 - 813.

Giesbrecht, F.G., Burns, J.C. (1985). Two-stage analysis based on a mixed model: Large-sample asymptotic theory and small-sample simulation results. *Biometrics* 41, 477 – 486.

Glaisher, J. W. L. (1873). On the Rejection of Discordant Observations. *Monthly Notices of the Royal Astronomical Society*, 23, 391-402.

Glaisher, J. W. L. (1874). Note on a paper by Mr. Stone "On the Rejection of Discordant Observations". *Monthly Notices of the Royal Astronomical Society*, 34, 251.

Gosset, W. S. (1908). The probable error of a mean. *Biometrika* 6 (1): 1–25 (published under the pen name "Student").

Gould, B. A. Jr. (1855). On Peirce's Criterion for the Rejection of Doubtful Observations, with Tables for Facilitating its Application. *Astronomical Journal*, 6, 81-83.

Harter, H. L. (1978). A chronological annotated bibliography of order statistics (Vol-ume I: pre—1950). *American Series in mathematical and management sciences*. Vol. 7. American Sciences Press, 515 pp.

Helmert, F. R. (1876). Ueber die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler und über einige damit im Zusammenhange stehende Fragen. *Zeitschrift für Mathematik und Physik*, 21, 102 - 219.

Henderson, C. R., O. Kempthorne, S. R. Searle and C. M. von Krosigk (1959). The Estimation of Environmental and Genetic Trends from Records Subject to Culling. *Biometrics* (International Biometric Society) 15 (2): 192–218.

Kackar, R.N., Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association* 79, 853 – 862.

Kenward, M. G. and Roger, J. H. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics* 53, 983-997.

Khuri, A. I., T. Mathew and B. K. Sinha (1998). *Statistical tests for Mixed Linear Models*. Wiley, New York.

Kok, J. J. (1984). On Data Snooping and Multiple Outlier Testing. *NOAA Technical Teport*, NOS NGS 30.

Krakiwsky, E. and Z. F. Biacs (1990): Least Squares Collocation and Statistical Testing. *Bull. Geod*. 64, 73-87.

McLean, R. A. and W. L. Sanders (1988). Approximated Degrees of Freedom for Standard Errors in Mixed Linear Models. In Proceedings of the Statistical Computing Section, *American Statistical Association*, 50-59.

Neyman, J. and E. Pearson (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: part I. *Biometrika* 20A, 175–240

Nobre, J. S. and J. M. Singer (2007). Residual analysis for linear mixed models. *Biometrical Journal* 49, 863-875.

Patterson, H. D. and R. Thompson (1971). Recovery of inter-block information when block sizes are unequal. *Biometrica* 58, 545-554.

Pearson, K. (1900). On the Criterion that a given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be reasonably supposed to have arisen from Random Sampling. *Philosophical Magazine Series* 5, 50 (302): 157 - 175.

Pearson, E. and C. Sekar (1936). The Efficiency of Statistical Tools and a Criterion for the Rejection of Outlying Observations. *Biometrica*, 28, 308-320.

Pope, A. J. (1976). The Statistics of Residuals and the Detection of Outliers. *NOAA Technical Report*, NOS 65, NGS 1.

Peirce, B. (1852). Criterion for the Rejection of Doubtful Observations. *Astronomical Journal*, 2, 161-163.

Rider, P. R. (1933). Criteria for Rejections of Observations. *Washington University Studies – New Series*, Science and Technology, 8.

Rossikopoulos, D. (1992). *Surveying Networks and Computations*. Editions Ziti (in Greek).

Rossikopoulos, D. (2015). Craps, meridians, genes and geographies. A brief history of statistics. In special issue for Emeritus Professor Myron Myridis: Arvanitis et al. (Eds.). *Cartographies of Mind, Soul and Knowledge*, Aristotle University of Thessaloniki, 134-152 (in Greek).

Rossikopoulos, D. (2017). Statistical testing for mixed linear models. In special issue in Honor of the Memory of Professor Ioannis Paraschakis: Fotiou et al. (Eds.). *Living with GIS*, Aristotle University of Thessaloniki, 9-40 (in Greek).

Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin* 2 (6), 110 – 114.

Saunders, S. A. (1903). Note on the Use of Peirce's Criterion for the Rejection of Doubtful Observations. *Monthly Notices of the Royal Astronomical Society*, 63, 432-436.

Schaffrin, B. (1987). Less Sensitive tests by introducing stochastic linear hypothesis. Proc. 2nd Inter. *Tampere Conference in Statistics*. Finland, June 1987.

Schaffrin, B. (1988). Tests for random effects based on homogeneously linear predictors. *Workshop on Theory and Practice in Data Analysis*, 19-21 August, Berlin.

Schaffrin, B. and Y. Bock (1994): Geodetic deformation analysis based on robust in-verse theory. *Manusscripta geodaetica*, 19, 31-44.

Seely, J. (1970). Linear spaces and unbiased estimation. *Ann. Math. Statist*. 41, 1725-1734.

Seely, J. (1970). Linear spaces and unbiased estimation – an application to the mixed linear model. *Ann. Math. Statist*. 41, 1735-1748.

Selvin, H. C. and A. Stuart (1966). Data dredging procedures in survey analysis. *The American Statistician*, 20(3), 20-23.

Singer, J. M., J. S. Nobre and F. M. Rocha (2013). Diagnostic and treatment for linear mixed models. *Proceedings, 59th ISI World Statistics Congress*, 25-30 August 2013, Hong Kong (Session CPS203), 5486-5491.

Srikantan, K. S. (1961). Testing for Single Outlier in a Regression Model. Sankhya, Ser. A, 23, 251-260.

Stone, E. J. (1868). On the Rejection of Discordant Observations. *Monthly Notices of the Royal Astronomical Society*, 28, 165-168.

Thompson, W. (1935). On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. *Annals of Mathematical Statistics*, 6.

You C., S. Müller and J. T. Ormerod (2016). On generalized degrees of freedom with application in linear mixed models selection. *Stat. Comput*. 26, 199-210.

Wald, A. (1943). Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. *Transactions of the American Mathematical Society*, 54.

Wei, M. (1987): Statistical Problems in Collocation. *Manuscripta Geodaetica*, 12, 282-289.

Weisberg, S. (1980). *Applied Linear Regression*. Wiley, New York.

Wintlock, J. (1856). On Professor Airy's Objections to Peirce's Criterion. *Astronomical Journal*, 4, 145-147.

Witkovsky V. (2012). Estimation, testing and prediction regions of the fixed and random effects by solving Henderson's mixed model equations. *Measurement Science Review*, Vol. 12, No. 6, 234-248.