

COMO ESTIMAR O PODER DO TESTE MÍNIMO E VALORES LIMITES PARA O INTERVALO DE CONFIANÇA DO DATA SNOOPING

How to estimate the minimum power of the test and bound values for the confidence interval of Data Snooping procedure

IVANDRO KLEIN^{1,2}
MARCELO TOMIO MATSUOKA^{1,2}
MATHEUS PEREIRA GUZATTO¹

¹Laboratório de Pesquisas em Geodésia (LAGEO)/Departamento de Geodésia/Instituto de Geociências – UFRGS

²Programa de Pós-Graduação em Sensoriamento Remoto – UFRGS
Av. Bento Gonçalves, 9500, CEP 91501-970,
Cx. Postal 15001, Porto Alegre, RS, Brasil
ivandroklein@gmail.com; tomiomatsuoka@gmail.com;
matheusguzatto@gmail.com

RESUMO

O *Data Snooping* (DS) é o método mais bem estabelecido para identificar erros grosseiros (*outliers*) em dados geodésicos, com uma determinada probabilidade. O poder do teste do DS é a probabilidade deste identificar corretamente um erro grosseiro, enquanto o intervalo de confiança do DS é a probabilidade deste não rejeitar uma observação não contaminada por erro grosseiro. Na prática, o poder do teste é sempre desconhecido. Desta forma, o objetivo deste trabalho é apresentar uma revisão teórica sobre como determinar o poder do teste mínimo e valores limites para o intervalo de confiança do DS, em um cenário n -dimensional, ou seja, considerando todas as observações envolvidas. Além da revisão teórica, um exemplo numérico envolvendo uma rede de nivelamento geométrico simulada é apresentado. Os resultados obtidos nos experimentos foram concordantes com os valores teóricos previamente calculados, ou seja, a metodologia revista apresentou desempenho satisfatório na prática. O exemplo apresentado também evidencia a

importância da metodologia revista na etapa de planejamento (ou pré-análise) de redes geodésicas.

Palavras-chave: Poder o Teste; Intervalo de Confiança; Data Snooping.

ABSTRACT

Data Snooping (DS) is the best-established method to identify gross errors (outliers) in geodetic data analysis, with a given probability. The power of the test is the probability of DS correctly identifying a gross error, while the confidence interval is the probability of DS not rejecting an observation uncontaminated by gross error. In practice, the power of the test is always unknown. Thus, the objective of this paper is to present a theoretical review of how to determine the minimum power of the test, and bound values for the confidence interval of the DS procedure in an n -dimensional scenario, i.e., considering all observations involved. Along with the theoretical review, a numerical example involving a simulated leveling network is presented. The results obtained in the experiments agreed with the previously calculated theoretical values, i.e., the revised methodology showed satisfactory performance in practice. The example also shows the importance of the revised methodology in the planning stage (or pre-analysis) of geodetic networks.

Keywords: Power of the Test; Confidence Interval; Data Snooping.

1. INTRODUÇÃO

O *Data Snooping* (DS) é o método mais estabelecido para identificar *outliers* em dados geodésicos (LEHMANN, 2012). O DS é um procedimento iterativo de testes estatísticos que identifica observações suspeitas de estarem contaminadas por erros grosseiros (*outliers*), inicialmente proposto em Baarda (1968), e que será descrito em detalhes na próxima seção.

Um teste estatístico consiste em dividir o espaço amostral (conjunto de todos os resultados possíveis de um experimento) em duas partes, denominadas região de rejeição (região crítica) e região de aceitação (ou de não rejeição) da hipótese de teste (ou hipótese nula). Se a estatística calculada, em função da amostra, se situar dentro da região crítica, rejeita-se a hipótese em questão. Caso contrário, não há evidências, na amostra observada, para rejeitar a hipótese nula (LARSON, 1974).

Portanto, quando o teste de uma hipótese estatística é realizado e uma decisão é tomada, pode-se cometer dois tipos de erro na decisão (LARSON, 1974):

- Erro Tipo I: Rejeitar a hipótese de teste quando esta é de fato verdadeira;
- Erro Tipo II: Não rejeitar a hipótese de teste quando esta de fato é falsa;

A probabilidade de cometer o Erro Tipo I é dada por α (nível de significância do teste), e a probabilidade de cometer o Erro Tipo II é dada por β . Designando a hipótese que se deseja testar por H_0 (hipótese nula), estas definições são formalmente definidas por (LARSON, 1974):

$$\begin{cases} \alpha = P(\text{erro tipo I}) = P(\text{rejeitar } H_0 | H_0 \text{ verdadeira}) \\ \beta = P(\text{erro tipo II}) = P(\text{não rejeitar } H_0 | H_0 \text{ falsa}) \end{cases}$$

No caso do procedimento DS, a hipótese nula (H_0) é a ausência de erros grosseiros nas observações, e para cada i -ésima observação do conjunto de dados, defini-se uma hipótese alternativa (H_i) como sendo a presença de erro grosseiro nesta i -ésima observação testada (BAARDA, 1968; TEUNISSEN, 2006). Ou seja, o DS é um procedimento de teste estatístico que possui múltiplas hipóteses alternativas a hipótese nula (no caso, existência de um único erro grosseiro para cada i -ésima observação testada individualmente).

Desta forma, por ser um procedimento de teste estatístico, e pelo fato que este considera somente uma observação suspeita de estar contaminada por erro grosseiro por vez (isto é, são múltiplas hipóteses alternativas a hipótese nula, uma para cada observação), o DS está sujeito a três tipos de erros (FÖRSTNER, 1983):

- Erro Tipo 1: Identificar (erroneamente) uma observação como contendo erro grosseiro (isto é, rejeitar H_0 quando H_0 é verdadeira);
- Erro Tipo 2: Não identificar (erroneamente) uma observação como contendo erro grosseiro (isto é, não rejeitar H_0 , ou em outras palavras, rejeitar H_i , quando H_0 é falsa, ou de maneira análoga, H_i é verdadeira);
- Erro Tipo 3: Identificar (erroneamente) uma outra observação como contendo erro grosseiro, enquanto a observação contaminada não foi identificada (isto é, aceitar uma hipótese alternativa H_j falsa, enquanto a hipótese alternativa verdadeira H_i não foi aceita).

Além desses três tipos de erros, tem-se ainda o intervalo de confiança (IC), que é a probabilidade de não rejeitar a hipótese nula (H_0), quando a hipótese nula é verdadeira, e o poder do teste (γ), que é a probabilidade de rejeitar a hipótese nula (ou em outras palavras, aceitar a hipótese alternativa), quando a hipótese nula é falsa (ou de maneira análoga, a hipótese alternativa é verdadeira) (LARSON, 1974).

Como no caso do DS, tem-se n hipóteses alternativas a hipótese nula (uma para cada observação testada), para cada observação, o intervalo de confiança (IC) se torna a probabilidade desta observação, caso não contaminada por erro grosseiro, não ser identificada como *outlier*, e o poder do teste (γ) se torna a probabilidade desta observação, caso contaminada por erro grosseiro, ser identificada como *outlier* (ver, por exemplo, BAARDA, 1968, KAVOURAS, 1982, TEUNISSEN, 2006).

O poder do teste (γ), naturalmente, é sempre desconhecido na prática, mas a sua relação com os resultados obtidos aplicando o DS foi estudada inicialmente em Baarda (1968), onde o mesmo autor propõe as chamadas medidas de confiabilidade para o DS, mas considerando apenas os efeitos do Erro Tipo 1 e do Erro Tipo 2 do teste (ou seja, desconsiderando a ocorrência do Erro Tipo 3).

Förstner (1983) estudou a relação do Erro Tipo 1, do Erro Tipo 2 e do Erro Tipo 3 no DS, mas considerando apenas duas observações por vez, adicionando na análise original de Baarda (1968) o coeficiente de correlação entre as estatísticas de teste de cada par de observações, ou seja, a chamada “análise de separabilidade”.

Considerando o fato de que o poder do teste é sempre desconhecido na prática, Yang et al. (2013), utilizando os estudos de Förstner (1983), propõem uma forma de se estimar o "poder do teste mínimo" do DS, mas nesta abordagem, considerando todas as observações envolvidas, ao invés de cada par de observações separadamente, além de definir valores máximos e mínimos para o intervalo de confiança do DS (também considerando todas as observações envolvidas).

A análise dos erros tipo 1, 2 e 3, e, consequentemente, do intervalo de confiança e do poder do teste, é muito importante na etapa de controle de qualidade de redes geodésicas, pois, considerando os custos do projeto e a qualidade (acurácia) final desejada para a rede, espera-se que apenas as observações de fato contaminadas por erros grosseiros sejam devidamente identificadas nesta etapa.

Além disso, na própria fase de planejamento ou pré-análise da rede, pode-se desejar que a mesma, em função da precisão e geometria/redundância das observações, tenha uma proteção razoável para os erros tipo 1, 2 e 3, ou seja, que a magnitude esperada para estes esteja dentro de uma determinada "margem de segurança" estipulada.

Desta forma, o objetivo deste trabalho é apresentar como se estima o poder do teste mínimo do procedimento DS, bem como valores limites para o intervalo de confiança deste, por meio da exposição teórica sobre o tema, além da aplicação e comprovação da metodologia em um exemplo numérico empregando o método Monte Carlo em uma rede de nivelamento geométrico.

É importante ressaltar que existem métodos alternativos à detecção de erros grosseiros em redes geodésicas por meio de testes estatísticos como o DS, dentre os quais se pode citar os métodos de estimação robusta (ver, por exemplo, KOCH, 1999), métodos de inferência bayesiana (ver, por exemplo, GUI et al., 2011), e o método QUAD (ver, por exemplo, GUO et al., 2007).

2. O PROCEDIMENTO DE TESTE *DATA SNOOPING*

No caso do procedimento de teste *Data Snooping*, proposto em Baarda (1968), inicialmente, considera-se o seguinte modelo linear de ajustamento:

$$Lb = Ax + e, \quad (1)$$

onde na Expressão 1, Lb é o vetor das n observações, x é o vetor dos u parâmetros incógnitos do modelo (no caso de redes geodésicas, as coordenadas dos vértices), A é a matriz *design* do ajustamento (ou matriz jacobiana), e e é o vetor dos n erros aleatórios das observações.

Para os casos em que o modelo matemático é inicialmente não linear em relação aos parâmetros, deve-se utilizar o modelo linearizado de Gauss-Markov (ver, por exemplo, GUO et al., 2007).

Na Expressão 1, assumindo que a esperança matemática dos erros aleatórios é igual ao vetor nulo ($E\{e\} = 0$), ou seja, por serem de natureza aleatória, estes erros não apresentam tendência, para cada i -ésima observação (lb_i) do vetor das n

observações (Lb), pode-se formular as seguintes hipóteses de teste (BAARDA, 1968; TEUNISSEN, 2006):

$$H_0: E\{Lb\} = Ax \quad \text{contra} \quad (2)$$

$$H_A: E\{Lb\} = Ax + c_i \nabla_i, \quad \nabla_i \neq 0,$$

onde na Expressão 2, c_i é um vetor unitário contendo a unidade na linha da i -ésima observação testada e zero nas demais linhas, ou seja: $c_i = \left[0 \quad 0 \quad \dots \quad 0 \quad \underset{i}{1} \quad 0 \quad 0 \quad \dots \quad 0 \right]_{nx1}^T$, e ∇_i é um (possível) erro grosseiro na i -ésima observação testada.

Desta forma, na hipótese nula (H_0), assume-se que não existe erro grosseiro nas observações, enquanto na hipótese alternativa (H_A), assume-se que a i -ésima observação testada (lb_i) está contaminada por um erro grosseiro de magnitude ∇_i .

A hipótese nula é rejeitada (ou em outras palavras, a hipótese alternativa é aceita), se a estatística de teste da i -ésima observação testada (w_i) exceder o seguinte valor crítico (BAARDA, 1968; TEUNISSEN, 2006):

$$w_i = \frac{c_i^T P v}{\sqrt{c_i^T P \Sigma_v P c_i}}; \quad |w_i| > c_{\alpha_0/2}, \quad (3)$$

onde na Expressão 3, P é a matriz peso do ajustamento, v é o vetor dos resíduos ajustados, Σ_v é a matriz de covariância dos resíduos ajustados, e $c_{\alpha_0/2}$ é o valor crítico teórico na distribuição normal padrão, para um dado nível de significância α_0 (teste bi-lateral).

Na prática, adota-se um nível de significância para o teste (como por exemplo, $\alpha_0 = 0,001$ ou $\alpha_0 = 0,01$), em função do qual se obtém o valor crítico teórico ($c_{\alpha_0/2}$), e testam-se todas as observações individualmente (ou seja, $i = 1, 2, 3, \dots, n$).

Como cada observação é testada individualmente, a (única) observação considerada contaminada por erro grosseiro será aquela cuja estatística de teste w_i satisfazer as seguintes condições (BAARDA, 1968; TEUNISSEN, 2006):

$$\begin{cases} |w_i| > c_{\alpha_0/2} \\ |w_i| > |w_j|, \quad \forall i \neq j \end{cases} \quad (4)$$

Identificada a observação suspeita de estar contaminada por erro grosseiro, segundo o nível de significância estipulado (α_0), usualmente, exclui-se a mesma do conjunto de dados e repete-se o ajustamento e o procedimento de teste DS até que todas as observações suspeitas sejam devidamente identificadas, em um processo

iterativo de ajustamento, identificação e remoção de erros grosseiros (um por vez).

Para mais detalhes sobre este procedimento, ver, por exemplo, Baarda (1968), Kavouras (1982), Koch (1999), Teunissen (2006) e Lehmann (2012).

3. NÍVEIS DE PROBABILIDADE ASSOCIADOS AO DATA SNOOPING CONSIDERANDO DUAS HIPÓTESES ALTERNATIVAS

Conforme já mencionado, ao se aplicar o procedimento de teste DS, podem ocorrer três tipos de decisões falsas: Identificar (erroneamente) uma observação como contendo erro grosseiro (Erro Tipo 1); não identificar (erroneamente) uma observação como contendo erro grosseiro (Erro Tipo 2); e identificar (erroneamente) uma observação como contendo erro grosseiro, enquanto a observação contaminada não foi identificada (Erro Tipo 3).

A probabilidade de cometer o Erro Tipo 1 é chamada de nível de significância do teste e é dada por α ; a probabilidade de cometer o Erro Tipo 2 é dada por β ; e a probabilidade de cometer o Erro Tipo 3 é dada por κ .

Em contrapartida, o intervalo de confiança (IC) é o complemento do Erro Tipo 1, ou seja, é a probabilidade de não identificar (corretamente) uma observação como contendo erro grosseiro ($IC = 1 - \alpha$), enquanto o poder do teste, designado por γ , é o complemento da soma das probabilidades do Erro Tipo 2 e do Erro Tipo 3 ($\gamma = 1 - (\beta + \kappa)$), ou seja, é a probabilidade de identificar (corretamente) uma observação como contendo erro grosseiro.

Portanto, o intervalo de confiança e o poder do teste são as probabilidades do resultado do teste conduzir a decisões corretas, ao contrário da ocorrência dos erros tipo 1, 2 e 3 (para mais detalhes, ver, por exemplo, FÖRSTNER, 1983; TEUNISSEN, 2006). Quando o valor para a probabilidade do Erro Tipo 1, do Erro Tipo 2 ou do poder do teste é pré-definido/estipulado, este valor é designado por α_0 , β_0 ou γ_0 , respectivamente.

Desta forma, considerando inicialmente o caso unidimensional envolvendo uma única observação testada e a sua estatística de teste w_i no DS (ou seja, desconsiderando a ocorrência do Erro Tipo 3), uma visualização do intervalo de confiança, do poder do teste, do nível de significância e do Erro Tipo 2 é obtida por meio da Figura 1 (neste caso, os valores assumidos para estes níveis de probabilidade são: $\alpha_0 = 0,001$; $IC = 1 - \alpha_0 = 0,999$; $\beta_0 = 0,20$; e $\gamma_0 = 1 - \beta_0 = 0,80$).

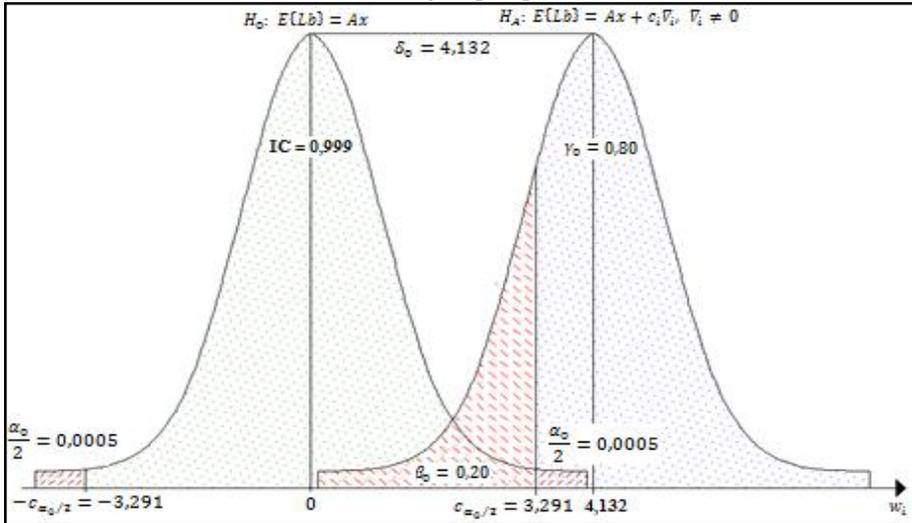
Analisando a Figura 1, nota-se que, aumentando o nível de significância (α_0) do DS, e conseqüentemente, diminuindo o valor crítico teórico ($c_{\alpha_0/2}$) e o intervalo de confiança (IC), aumenta-se o poder do teste (γ_0), ou analogamente, diminui-se a probabilidade do Erro Tipo 2 (β_0).

Em contrapartida, diminuindo o nível de significância do DS, e conseqüentemente, aumentando o valor crítico teórico e o intervalo de confiança, diminui-se o poder do teste, ou analogamente, aumenta-se a probabilidade do Erro Tipo 2. Ou seja, na prática, não é possível minimizar, simultaneamente, as

probabilidades de cometer o Erro Tipo 1 e o Erro Tipo 2 (ou de maneira análoga, maximizar, simultaneamente, o intervalo de confiança e o poder do teste).

Além disso, analisando ainda a Figura 1, nota-se que a probabilidade do Erro Tipo 2 (β_0), e consequentemente, o poder do teste (γ_0), dependem, além do valor crítico teórico ($c_{\alpha_0/2}$), da separação entre a hipótese nula (ausência de erros grosseiros nas observações – H_0) e a hipótese alternativa (presença de erro grosseiro na i -ésima observação testada – H_A), ou seja, do parâmetro de não centralidade correspondente do modelo (δ_0).

Figura 1 – Níveis de probabilidade associados ao DS no teste de uma i -ésima observação qualquer.



Como a magnitude do (possível) erro grosseiro que contamina uma i -ésima observação testada é sempre desconhecida, na prática, arbitra-se um valor para o nível de significância (α_0) e para o poder do teste (γ_0), em função dos quais se obtém o valor para o parâmetro de não centralidade do modelo correspondente (δ_0). Para mais detalhes sobre esta abordagem, ver, por exemplo, Baarda (1968), Kavouras (1982), Koch (1999) e Teunissen (2006).

Considerando agora um caso bidimensional envolvendo duas observações quaisquer, cujas estatísticas de teste sejam dadas por w_i e w_j (ou seja, adicionando a probabilidade do Erro Tipo 3 no cenário unidimensional descrito anteriormente), todas as situações (e decisões) relativas ao DS podem ser reunidas de acordo com a Tabela 1.

Na Tabela 1, H_0 é a hipótese nula do DS (não existência de erros grosseiros nas observações), H_i é a hipótese alternativa para a i -ésima observação (existência

de erro grosseiro nesta), e H_j é a hipótese alternativa para a j -ésima observação (existência de erro grosseiro nesta).

Tabela 1 – Decisões associadas a duas hipóteses alternativas no DS
(Fonte: Adaptado de Yang *et al.*, 2013).

Hipótese “aceita” no teste	Resultado do teste		
	H_0	H_i	H_j
Critério de decisão	$ w_i \leq c_{\alpha 0/2},$ $ w_j \leq c_{\alpha 0/2}$	$ w_i > c_{\alpha 0/2},$ $ w_i > w_j $	$ w_j > c_{\alpha 0/2},$ $ w_j > w_i $
Realidade (“desconhecida”)			
Hipótese nula (H_0)	Decisão correta	Erro Tipo 1	Erro Tipo 1
Probabilidade associada	$IC = 1 - \alpha_{00}$	α_{oi}	α_{oj}
Hipótese alternativa (H_i)	Erro Tipo 2	Decisão correta	Erro Tipo 3
Probabilidade associada	β_{i0}	$\gamma_{ii} = 1 - \beta_{ii}$	κ_{ij}
Hipótese alternativa (H_j)	Erro Tipo 2	Erro Tipo 3	Decisão correta
Probabilidade associada	β_{j0}	κ_{ji}	$\gamma_{jj} = 1 - \beta_{jj}$

Desta forma, α_{00} é o nível de significância “global” do DS (enquanto α_{oi} é o nível de significância apenas para o teste da i -ésima observação, e α_{oj} é o nível de significância apenas para o teste da j -ésima observação).

Além disso, β_{i0} e κ_{ij} são as probabilidades de cometer, respectivamente, o Erro Tipo 2 e o Erro Tipo 3 para a i -ésima observação, enquanto β_{j0} e κ_{ji} são as probabilidades de cometer, respectivamente, o Erro Tipo 2 e o Erro Tipo 3 para a j -ésima observação.

Finalmente, β_{ii} e γ_{ii} são, respectivamente, a soma de β_{i0} e κ_{ij} e o poder do teste para a i -ésima observação, enquanto β_{jj} e γ_{jj} são, respectivamente, a soma de β_{j0} e κ_{ji} e o poder do teste para a j -ésima observação.

As estatísticas de teste w_i e w_j de duas observações quaisquer possuem um coeficiente de correlação (ρ_{ij}), dado pela seguinte expressão (FÖRSTNER, 1983):

$$\rho_{ij} = \frac{c_i^T P \Sigma_v P c_j}{\sqrt{c_i^T P \Sigma_v P c_i} \sqrt{c_j^T P \Sigma_v P c_j}}; \quad -1 \leq \rho_{ij} \leq 1 \quad (5)$$

Além do coeficiente de correlação entre duas estatísticas de teste do DS, tem-se ainda, conforme visto anteriormente, o parâmetro de não centralidade do modelo (δ_0), que expressa a separação entre a hipótese nula e a hipótese alternativa.

No caso bidimensional aqui considerado, o parâmetro de não centralidade do modelo (δ_0) depende do nível de significância do teste (α_0), ou analogamente, do valor crítico teórico para este ($c_{\alpha_0/2}$), do poder do teste (γ_0), ou analogamente, da soma das probabilidades do Erro Tipo 2 (β_{0i}) e do Erro Tipo 3 (κ_{ij}), e também do coeficiente de correlação entre as estatísticas de teste consideradas (ρ_{ij}).

Desta forma, matematicamente, as probabilidades de cometer o Erro Tipo 1 (α_{0i}), o Erro Tipo 2 (β_{0i}) e o Erro Tipo 3 (κ_{ij}) no teste de uma i -ésima observação qualquer, considerando o coeficiente de correlação (ρ_{ij}) da estatística de teste desta (w_i) com a de uma j -ésima observação qualquer (w_j), bem como o parâmetro de não centralidade do modelo correspondente (δ_0), são dadas, respectivamente, por (FÖRSTNER, 1983):

$$\alpha_{0i} = \iint_{|w_i| > c_{\alpha_0/2}, |w_i| > |w_j|} \frac{1}{2\pi\sqrt{1-\rho_{ij}^2}} e^{-\frac{1}{2(1-\rho_{ij}^2)}[w_i^2 - 2\rho_{ij}w_iw_j + w_j^2]} dw_i dw_j \quad (6)$$

$$\beta_{0i} = \iint_{|w_i| \leq c_{\alpha_0/2}, |w_j| \leq c_{\alpha_0/2}} \frac{1}{2\pi\sqrt{1-\rho_{ij}^2}} e^{-\frac{1}{2(1-\rho_{ij}^2)}[(w_i-\mu_i)^2 - 2\rho_{ij}(w_i-\mu_i)(w_j-\mu_j) + (w_j-\mu_j)^2]} dw_i dw_j \quad (7)$$

$$\kappa_{ij} = \iint_{|w_j| > c_{\alpha_0/2}, |w_j| > |w_i|} \frac{1}{2\pi\sqrt{1-\rho_{ij}^2}} e^{-\frac{1}{2(1-\rho_{ij}^2)}[(w_i-\mu_i)^2 - 2\rho_{ij}(w_i-\mu_i)(w_j-\mu_j) + (w_j-\mu_j)^2]} dw_i dw_j \quad (8)$$

Na Expressão 6, o termo dentro da integral corresponde à função densidade de probabilidade (fdp) conjunta de w_i e w_j na hipótese nula H_0 (ausência de erros grosseiros nas duas observações), enquanto nas expressões 7 e 8, o termo dentro das integrais corresponde a fdp conjunta de w_i e w_j na hipótese alternativa H_i (existência de erro grosseiro na i -ésima observação, com $\mu_i = \delta_0$ e $\mu_j = \rho_{ij} \delta_0$).

Como as distribuições de probabilidade de w_i e w_j em H_i e H_j são simétricas, devido ao coeficiente de correlação entre estas, decorre que $\alpha_{0i} = \alpha_{0j}$, $\beta_{0i} = \beta_{0j}$, e $\kappa_{ij} = \kappa_{ji}$ (para mais detalhes, ver FÖRSTNER, 1983 e YANG et al., 2013).

4. LIMITES PARA O INTERVALO DE CONFIANÇA E PARA O PODER DO TESTE DO DATA SNOOPING NO CENÁRIO n -DIMENSIONAL

Seguindo a formulação teórica descrita anteriormente, por meio das distribuições de probabilidade resultantes, pode-se calcular o Erro Tipo 1, o Erro

Tipo 2, o Erro Tipo 3, o intervalo de confiança e o poder do teste do DS em um cenário tridimensional, tetradimensional, e até mesmo, estendendo até o caso n -dimensional, ou seja, considerando todas as n observações testadas (individualmente).

Entretanto, isto envolveria integrações numéricas n -dimensionais, o que impossibilita o seu cálculo na prática.

Desta forma, Yang et al. (2013), utilizando a abordagem bi-dimensional apresentada anteriormente, definem limites para o intervalo de confiança e para o poder do teste do DS em um cenário (geral) n -dimensional, ou seja, considerando todas as observações testadas.

Para o intervalo de confiança do DS, no cenário n -dimensional ($IC = 1 - \alpha_0$, sendo α_0 o nível de significância “global” do DS, ou seja, neste cenário n -dimensional), o IC terá como limites superior e inferior os seguintes valores (YANG et al., 2013):

$$1 - \alpha_0'' > 1 - \alpha_0 \geq (1 - \alpha_0)^n, \quad (9)$$

onde na Expressão 9, α_0'' corresponde ao nível de significância relativo ao par de estatísticas de teste com maior coeficiente de correlação (em módulo), obtido por meio da Expressão 6, n corresponde ao número total de observações, e α_0 corresponde ao nível de significância que foi arbitrado ou definido para o teste (como por exemplo, $\alpha_0 = 0,001$ ou $\alpha_0 = 0,01$), em função do qual se obtém o valor crítico teórico ($c_{\alpha_0/2}$) na distribuição normal padrão (teste bilateral).

Finalmente, para cada observação (no caso, para uma i -ésima observação qualquer), o poder do teste do DS ($\gamma_{ii} = 1 - \beta_{ii}$), no cenário n -dimensional, ou seja, considerando todas as demais observações testadas, terá como limite inferior o seguinte valor (YANG et al., 2013):

$$\gamma_{ii} = 1 - \beta_{ii} \geq 1 - (\beta_{i0}'' + \sum_{j=1, j \neq i}^n \kappa_{ij}), \quad (10)$$

onde na Expressão 10, β_{i0}'' corresponde a probabilidade do Erro Tipo 2 para a observação cuja estatística de teste apresenta o maior coeficiente de correlação (em módulo) com a estatística de teste da i -ésima observação considerada (obtida por meio da Expressão 7), e o somatório das probabilidades do Erro Tipo 3 ($\sum \kappa_{ij}$) deve ser calculado relacionando todas as demais observações (par a par) com a i -ésima observação considerada, por meio da Expressão 8.

Desta forma, o poder do teste mínimo do DS, no cenário n -dimensional (considerando todas as observações testadas), será igual ao poder do teste mínimo obtido por meio da Expressão 10, aplicada para cada observação individualmente, enquanto o intervalo de confiança mínimo do DS, neste cenário, será igual ao limite inferior dado pela Expressão 9.

Sobre estes níveis de probabilidade, é importante notar ainda que: Quanto

maior o coeficiente de correlação entre duas estatísticas de teste, maior é o tamanho do Erro Tipo 3 e menor é o tamanho do Erro Tipo 2, bem como, maior é o valor do parâmetro de não centralidade do modelo correspondente (para mais detalhes, ver, por exemplo, FÖRSTNER, 1983; YANG et al., 2013).

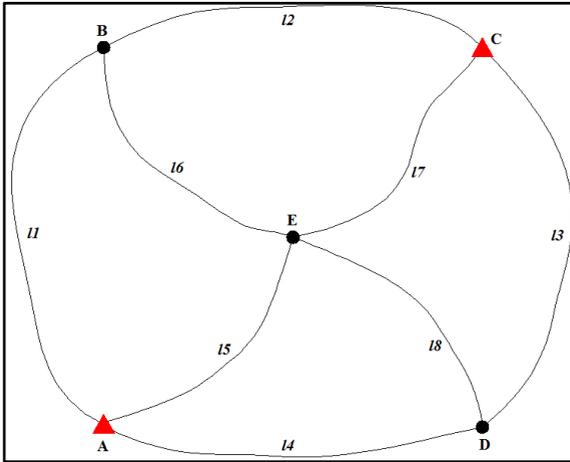
Ou seja, quanto maior o coeficiente de correlação entre duas estatísticas de teste, maior deve ser o parâmetro de não centralidade associado, e consequentemente, menor é a “separabilidade” entre estas estatísticas de teste, isto é, maior é o tamanho do Erro Tipo 3 para este par de observações.

Além disso, quanto maior o coeficiente de correlação máximo e o número total de observações (n), maior é a discrepância entre os limites dados pela Expressão 9 e o “verdadeiro” intervalo de confiança do DS no cenário n -dimensional (ver, por exemplo, LEHMANN, 2012).

5. EXEMPLO NUMÉRICO

Para mostrar na prática a determinação do intervalo de confiança e do poder do teste mínimos do DS, um exemplo numérico envolvendo uma rede de nivelamento geométrico simulada é apresentado. A rede consiste de $n = 8$ desníveis (observações), três vértices desconhecidos (B, D, e E) e dois conhecidos (A e C), conforme ilustra a Figura 2.

Figura 2 – Rede de nivelamento geométrico simulada.



Na rede numericamente simulada, foi considerado que todos os desníveis (observações) apresentam um desvio-padrão de $\sigma_1 = \pm 3$ mm, e que as cotas dos vértices A e C são fixas (injunções absolutas), e iguais a 100 m. As cotas dos vértices B e D, “isentas” de erros, são iguais a 105 m, e a cota do vértice E, “isenta” de erro, é igual a 110 m. Desta forma, foram aleatoriamente geradas as oito observações (considerando o desvio-padrão de ± 3 mm destas), e então foi

realizado o ajustamento pelo método dos mínimos quadrados desta rede de nivelamento.

Com as matrizes do ajustamento, por meio da Expressão (4), foram calculados os coeficientes de correlação (ρ_{ij}) das estatísticas de teste dos oito desníveis, conforme apresenta a Tabela 2.

Tabela 2 – Coeficientes de correlação (ρ_{ij}) entre as estatísticas de teste dos oito desníveis da rede.

Desnível	1	2	3	4	5	6	7
1							
2	0,5789						
3	-0,0526	0,0526					
4	0,0526	-0,0526	0,5789				
5	-0,1502	0,1502	-0,1502	0,1502			
6	0,4588	-0,4588	-0,1147	0,1147	-0,3273		
7	-0,1502	0,1502	-0,1502	0,1502	-0,4286	-0,3273	
8	-0,1147	0,1147	0,4588	-0,4588	-0,3273	-0,2500	-0,3273

Analisando a Tabela 2, nota-se que os coeficientes de correlação entre as estatísticas de teste das observações desta rede não são muito elevados, variando, em módulo, entre 0,05 e 0,58.

Para o cálculo do poder do teste mínimo e os limites do intervalo de confiança do DS no cenário n -dimensional, inicialmente, se calculou o valor do parâmetro de não centralidade do modelo correspondente para cada par de observações.

Para isto, o poder do teste de uma observação, no cenário bi-dimensional, foi fixado em $\gamma_{ii} = 0,80$. Desta forma, o parâmetro de não centralidade do modelo (δ_0) pode ser obtido por meio da seguinte expressão para o poder do teste de uma observação no cenário bidimensional (FÖRSTNER, 1983):

$$\gamma_{ii} = 1 - \beta_{ii} = \iint_{|w_i| > c_{\alpha/2}, |w_i| > |w_j|} \frac{1}{2\pi \sqrt{1-\rho_{ij}^2}} e^{-\frac{1}{2(1-\rho_{ij}^2)}[(w_i-\mu_i)^2 - 2\rho_{ij}(w_i-\mu_i)(w_j-\mu_j) + (w_j-\mu_j)^2]} dw_i dw_j, \tag{11}$$

onde na Expressão (11), tal como nas expressões (6) e (7), $\mu_i = \delta_0$ e $\mu_j = \rho_{ij} \delta_0$. Ou seja, fixando o resultado da integral em $\gamma_{ii} = 0,80$, inicia-se um processo iterativo de cálculo do parâmetro de não centralidade do modelo (δ_0) até que esta igualdade seja satisfeita.

No caso, o parâmetro de não centralidade do modelo foi inicialmente arbitrado em $\delta_0 = 3,231$, e, para cada par de observações, o poder do teste foi calculado por meio da Expressão (11). Para um resultado diferente de $\gamma_{ii} = 0,8$, o parâmetro de não

centralidade do modelo era aumentado em +0,0001, e o cálculo realizado novamente, até que a diferença entre o valor calculado por meio da Expressão (10) e o valor de referência ($\gamma_{ii} = 0,8$) do poder do teste ficasse na terceira casa decimal, o que corresponde a um erro máximo de 0,1% no resultado da integral.

O valor crítico adotado nestes casos foi de $c_{\alpha_0/2} = 2,5758293$, o que corresponde a um nível de significância estipulado em $\alpha_0 = 0,01$ (teste bilateral). É importante ressaltar que foram utilizadas sete casas decimais no valor crítico ($c_{\alpha_0/2}$) visando minimizar o erro de cálculo das integrações numéricas envolvidas na Expressão 10.

Desta forma, a Tabela 3 apresenta os valores obtidos para o parâmetro de não centralidade do modelo seguindo esta abordagem, para cada par de observações considerado.

Tabela 3 – Valores para o parâmetro de não centralidade do modelo (δ_0) para cada par de desníveis considerados (cenário bi-dimensional).

Par de desníveis considerados	1	2	3	4	5	6	7
1							
2	3,5139						
3	3,4211	3,4211					
4	3,4211	3,4211	3,5139				
5	3,4211	3,4211	3,4211	3,4211			
6	3,4642	3,4642	3,4211	3,4211	3,4366		
7	3,4211	3,4211	3,4211	3,4211	3,456	3,4366	
8	3,4211	3,4211	3,4642	3,4642	3,4366	3,4277	3,4366

Analisando as Tabelas 2 e 3, nota-se que, conforme já mencionado, quanto maior o coeficiente de correlação entre as estatísticas de teste de duas observações, maior é o valor do parâmetro de não centralidade do modelo correspondente. Para o cálculo numérico destas integrais bidimensionais, foi utilizada a função “*int2d*” do *software* gratuito *Scilab v. 5.4.1*.

Com os coeficientes de correlação entre as estatísticas de teste (Tabela 2) e os valores para o parâmetro de não centralidade do modelo obtidos no caso bidimensional (Tabela 3), foi calculado o poder do teste mínimo para cada observação, por meio da Expressão (9), e os resultados obtidos são apresentados na Tabela 4.

Analisando a Tabela 4, nota-se que os poderes de teste mínimos de cada observação não diferem muito entre si, o que pode ser explicado pelos coeficientes de correlação relativamente baixos entre as estatísticas de teste destas para a maior parte dos casos, além da própria geometria/configuração da rede de nivelamento simulada (Figura 2).

Por exemplo, os desníveis 1, 2, 3 e 4 envolvem um vértice desconhecido periférico e um vértice conhecido (apresentando a mesma estimativa para o poder do teste mínimo), os desníveis 6 e 8 envolvem dois vértices desconhecidos (apresentando as menores estimativas para o poder do teste mínimo), e os desníveis 5 e 7 envolvem o vértice desconhecido de maior redundância na rede (ponto central E) e um vértice conhecido, apresentando as maiores estimativas para o poder do teste mínimo.

Tabela 4 – Poder do teste mínimo de cada desnível no cenário n -dimensional.

Observação (desnível)	Poder do teste mínimo (em %)
1	76,39
2	76,39
3	76,39
4	76,39
5	76,62
6	74,80
7	76,62
8	74,80

Finalmente, por meio da Expressão (8), se calculou os limites para o intervalo de confiança do DS no cenário n -dimensional ($IC = 1 - \alpha_{00}$), resultando em:

$$0,9814 > 1 - \alpha_{00} \geq 0,9227$$

Analisando o resultado acima, nota-se que o intervalo de confiança do DS, no cenário n -dimensional, se situa entre 0,9227 e 0,9814, ou seja, valores estes menores do que o IC estipulado em $IC = 1 - \alpha_0 = 0,99$ na abordagem unidimensional.

Para verificar se de fato, os valores correspondentes para o poder do teste mínimo do DS obtidos com esta abordagem são realistas, simulações pelo método Monte Carlo foram realizadas na rede de nivelamento geométrico.

No caso, para cada observação, foram gerados e inseridos erros grosseiros (*outliers*), de magnitude entre 3σ e 9σ , sendo “ σ ” o desvio-padrão das observações, e foi aplicado o procedimento de teste DS com um nível de significância estipulado em $\alpha_0 = 1\%$, ou seja, com um valor crítico tabelado correspondente de $c_{\alpha_0/2} = 2,5758293$.

Para cada observação, 10.000 erros grosseiros foram gerados e inseridos nos experimentos, e os resultados obtidos são apresentados na Tabela 5.

Analisando e comparando as Tabelas 4 e 5, nota-se que de fato, para esta rede, todas as observações apresentaram um poder de teste superior ao valor mínimo estimado. Nota-se também que, para todas as observações, o poder do teste é relativamente alto, devido à geometria/configuração da rede simulada, com oito desníveis para apenas três vértices desconhecidos, ou seja, um ajustamento com

cinco graus de liberdade, e também pelos coeficientes de correlação relativamente baixos entre as estatísticas de teste das observações.

Para redes geodésicas com menor redundância, ou seja, com uma geometria mais “fraca”, e/ou cujas estatísticas de teste apresentem maior correlação, o poder do teste mínimo do DS seria menor.

Desta forma, este exemplo também ilustra a importância de se determinar o poder do teste mínimo logo na etapa de pré-análise ou planejamento da rede.

Tabela 5 – Resultados das 10.000 simulações para cada observação.

Observação	Nº de outliers detectados corretamente	Poder do teste (em %)
1	8.608	86,08
2	9.109	91,09
3	8.878	88,78
4	9.509	95,09
5	9.139	91,39
6	7.889	78,89
7	8.844	88,44
8	8.775	87,75

Neste caso, pode-se dizer que a rede de nivelamento geométrico em questão, com esta geometria/configuração, é resistente a um erro grosseiro (*outlier*). Se os valores obtidos para o poder do teste mínimos fossem muito mais baixos, a rede deveria ser “melhorada”, como por exemplo, com a inclusão de novas observações e/ou novos pontos de controle, até que os resultados obtidos fossem concordantes com valores previamente estipulados.

É importante ressaltar que os resultados obtidos na Tabela 4 são relativos aos casos em que o poder do teste de cada observação, no cenário bidimensional, é igual a $\gamma_{ii} = 0,80$, ou seja, a soma do Erro Tipo 2 e do Erro Tipo 3 é igual a $\beta_{ii} = 0,20$, pois foi este o valor de referência do resultado da integral na Expressão (10) para o cálculo dos valores do parâmetro de não centralidade do modelo correspondente apresentados na Tabela 3.

De qualquer forma, esta abordagem, mesmo “arbitrando” o poder do teste no cenário bi-dimensional, apresenta vantagens em relação à abordagem clássica apresentada em Baarda (1968), pois, no referido trabalho, desconsidera-se a ocorrência do Erro Tipo 3, e o poder do teste e as medidas de confiabilidade resultantes são baseados em valores arbitrados para uma abordagem unidimensional, ou seja, para cada observação separadamente.

Na metodologia aqui descrita, apesar dos valores arbitrados para o caso bidimensional, os resultados obtidos consideram a influência do Erro Tipo 3 e ainda são estendidos a um cenário n -dimensional, onde todas as observações são testadas (individualmente), sendo, portanto, resultados “menos otimistas” e “mais concordantes” com a realidade.

Em contrapartida, a principal desvantagem desta metodologia é o número de cálculos necessários, muito maior do que na abordagem convencional inicialmente proposta em Baarda (1968).

6. CONCLUSÃO

Neste trabalho, foi apresentada uma revisão teórica sobre como determinar o poder do teste mínimo e valores limites para o intervalo de confiança do procedimento DS, em um cenário n -dimensional, ou seja, considerando todas as observações envolvidas.

Para demonstrar na prática a metodologia revista, um exemplo numérico envolvendo uma rede de nivelamento geométrico simulada foi apresentado. Os resultados obtidos nos experimentos foram concordantes com os valores teóricos previamente calculados; isto mostra que a metodologia revista apresentou desempenho satisfatório na prática, apesar do grande número de cálculos necessários.

Como sugestões para trabalhos futuros, pode-se citar a aplicação da metodologia revista não apenas para o DS, mas também para os testes estatísticos para a identificação de múltiplos *outliers* (simultâneos) nas observações, a aplicação da metodologia revista nas chamadas medidas de confiabilidade dos testes, e também a sua aplicação ainda na etapa de pré-análise ou planejamento de redes geodésicas.

AGRADECIMENTOS

Os autores agradecem ao CNPq pela Bolsa de Produtividade em Pesquisa (Proc.n.303306/2012-2) do segundo autor e pelo projeto Universal (Proc. n. 477914/2012-8).

REFERÊNCIAS BIBLIOGRÁFICAS

- BAARDA, W. *A testing procedure for use in geodetic networks*. Publications on Geodesy, New Series, v. 2, n. 5, Delft: Netherlands Geodetic Commission, 1968.
- FÖRSTNER, W. Reliability and discernability of extended Gauss-Markov models. In: *Seminar on mathematical models to outliers and systematic errors*, Deutsche Geodätische Kommission, Series A, no. 98. Munich, Germany, pp 79–103, 1983.
- GUI, Q.; LI, X.; GONG, Y.; LI, B.; LI, G. A Bayesian unmasking method for locating multiple gross errors based on posterior probabilities of classification variables. *Journal of Geodesy*, v. 85, n. 4, p. 191–203, 2011.
- KAVOURAS, M. *On the Detection of Outliers and the Determination of Reliability in Geodetic Networks*. 1982. M.Sc.E. Thesis – Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, Canada, 1982.

- GUO, J. F.; OU, J. K.; WANG, H. Quasi-accurate detection of outliers for correlated observations. *Journal of Surveying Engineering*, v. 133, n. 3, p. 129–133, 2007.
- KOCH, K. R. *Parameter estimation and hypothesis testing in linear models*. 2. ed. Berlin: Springer, 1999.
- LARSON, H. J. *Introduction to probability theory and statistical inference*. 2. ed. New York: John Wiley & Sons, 1974. 430p.
- LEHMANN, R. Improved critical values for extreme normalized and studentized residuals in Gauss–Markov models. *Journal of Geodesy*, v. 86, n. 12, p. 1137–1146, 2012.
- TEUNISSEN, P. J. G. *Testing theory: an introduction*. 2. ed. Delft: Ed. VSSD, 2006. 147p.
- YANG, L.; WANG, J.; KNIGHT, N.; SHEN, Y. Outlier separability analysis with a multiple alternative hypotheses test. *Journal of Geodesy*, v. 85, n. 6, p. 591–604, 2013.

(Recebido em maio de 2014. Aceito em outubro de 2014).