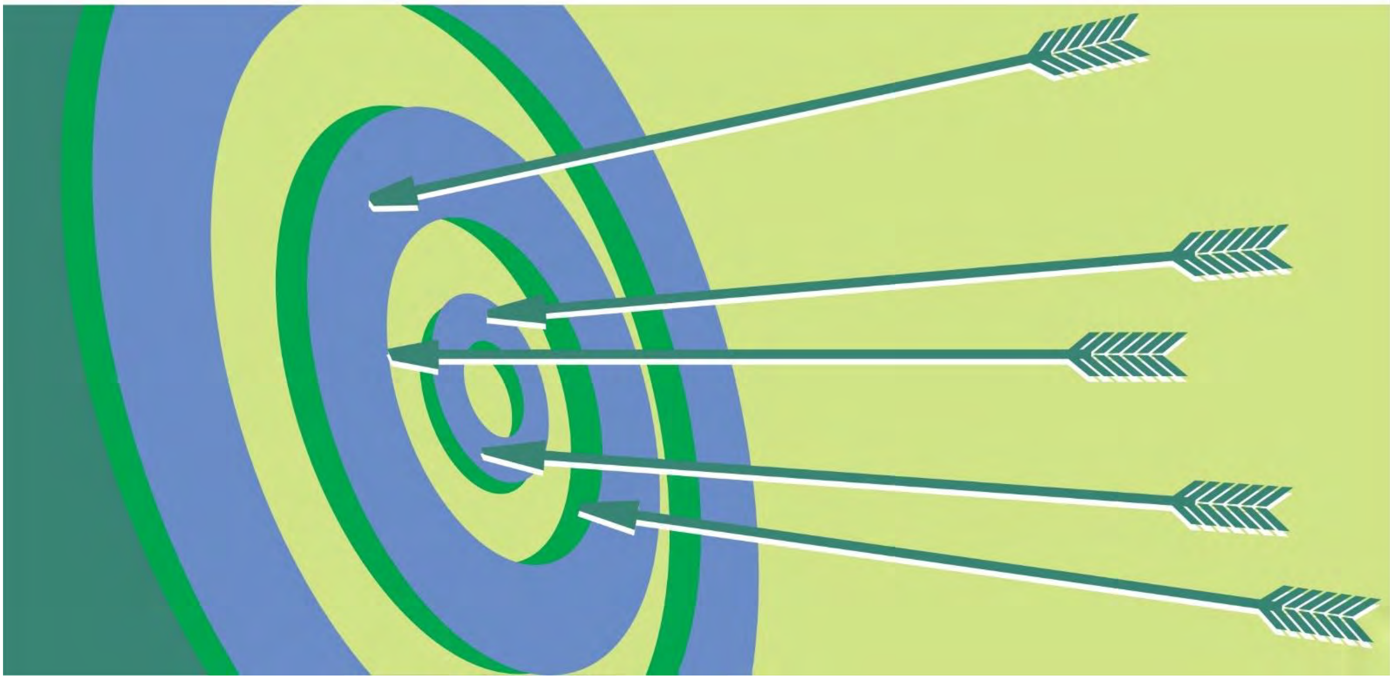
# Nota Técnica 269

## Acurácia e Precisão de Sistemas de Medição

Ming, Sun H.
Anselmo, Márcio A.
Carmo, José Antônio D. P.
Duarte, Tadeu L.
Fernandes, Marcelo A. F.
Lopes, Denise L.
Nahas, Caio R. F.
Santos, Alexandre F.
Souza, Cláudio P. A.
Tarricone, Nílvio A.
Ueta, Paulo S.



# 1. INTRODUÇÃO

O presente trabalho tem por objetivo fazer uma descrição do que significam a acurácia, a precisão e erro amostral de medições de tráfego. Por medições de tráfego entende-se as medidas de velocidade de veículos ou de contagem de veículos motorizados, bicicletas, pedestres ou de qualquer outra grandeza de tráfego. Existem atualmente disponíveis no mercado várias tecnologias que fazem essas medições. Mas, por melhor que seja a tecnologia, sempre existe um erro nessas medições. Nenhuma medição de qualquer grandeza física é exata. Note-se que o valor verdadeiro (*ground truth*) jamais é conhecido. As medidas obtidas são apenas aproximações do valor verdadeiro.

O objetivo é fazer com que o resultado medido seja o mais próximo possível do valor verdadeiro.
O trabalho também pretende apresentar um breve resumo das diversas metodologias descritas na literatura para aferir o erro dos sistemas de medição.

Para efeitos do presente trabalho, denota-se por "sistema de medição" o conjunto "equipamento (*hardware/software*) + procedimento de medição", uma vez que o resultado da medição depende não apenas da qualidade ou sofisticação do equipamento e do seu *software*, mas também da habilidade de quem está fazendo a medição e da correção do método adotado (instalação, configuração, calibração e ajustes).

A motivação do trabalho surgiu da necessidade de testar equipamentos que contam veículos, pedestres ou bicicletas, em substituição às contagens manuais. Desses testes adveio o problema de como aferir o erro desses equipamentos.

# 2. ACURÁCIA E PRECISÃO

Muitas vezes confunde-se a acurácia com a precisão. Na verdade, acurácia e precisão são dois conceitos distintos. O presente trabalho tenta esclarecer a diferença entre a acurácia e a precisão, bem como a sua relação com o erro das medições.

Acurácia e precisão, além de distintos, são conceitos independentes, entretanto, complementares.

## 2.1 Acurácia

A acurácia ou exatidão de um sistema de medição indica quão próximo está o resultado da medida do valor verdadeiro. A acurácia indica o grau de concordância que há entre o resultado da medição e o valor verdadeiro.

Por exemplo, considere dois sistemas de contagem de pedestres A e B. Suponha que o valor verdadeiro seja 100 pedestres e que o sistema A contou 110 pedestres e o sistema B contou 80 pedestres. Então, pode-se dizer que o sistema A tem uma acurácia maior do que o sistema B, pois o sistema A apresentou um desvio de 10% em relação ao valor verdadeiro, enquanto o sistema B teve um desvio de 20%.

A acurácia está relacionada com os erros aleatórios e com os erros sistemáticos.

Segundo o artigo "Erros aleatórios e sistemáticos: O que é isso?" (Fonte: https://accmetrologia.com.br/erros-aleatorios-e-sistematico-o-que-e-isso/):

> *"O erro aleatório é a parcela imprevisível do erro e se origina de variações temporais ou espaciais. Diferente do erro sistemático, o erro aleatório não pode ser eliminado nem corrigido, mas geralmente reduzido".*

Os erros sistemáticos resultam de fatores ligados às limitações, deficiências e/ou falhas do sistema de medição. Falhas e deficiências do equipamento e do seu *software* ocasionam erros sistemáticos de medição. Falhas e erros de configuração e calibração ou equipamentos mal ajustados também resultam em erros sistemáticos. Os erros sistemáticos acarretam uma pobre acurácia e conduzem a resultados falsos e enviesados, devendo ser corrigidos.

## 2.2 Precisão

A precisão de um sistema de medição indica o grau de variação de um conjunto de medições. Precisão é o grau de concordância entre vários resultados de medição obtidos com o mesmo procedimento e método.

Por exemplo, considere que foram realizadas 5 contagens de pedestres usando os sistemas A e B, com os seguintes resultados, supondo que o número de pedestres nos 5 intervalos em que foram feitas as contagens seja constante (ver a Tabela 1):

Tabela 1 - Dispersão dos valores medidos

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| A | 20 | 30 | 10 | 25 | 15 |
| B | 25 | 24 | 26 | 25 | 27 |

Pela menor dispersão ou espalhamento dos valores apresentados, pode-se dizer que o sistema B é mais preciso que o sistema A. Note que o conceito de precisão não depende do valor verdadeiro.
A precisão está relacionada apenas a erros aleatórios.
O ideal é que um sistema de contagem tenha boa acurácia e boa precisão.
A Figura 1 abaixo ilustra os conceitos de acurácia e precisão.

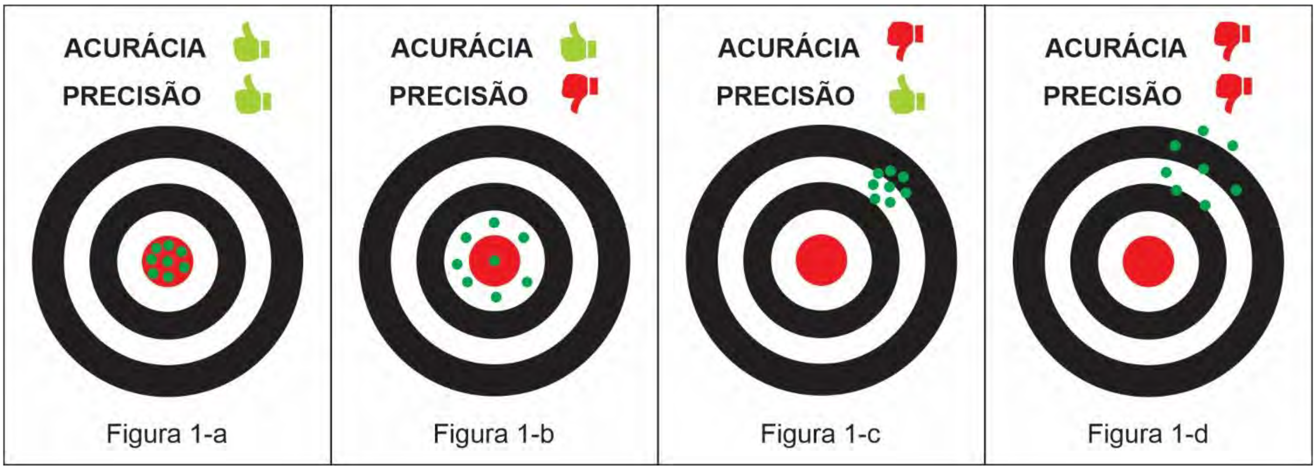



*Figura 1 - Ilustração dos conceitos de acurácia e precisão*
*Fonte: https://blog.idwall.co/o-que-e-acuracia/*

Na Figura 1, o círculo vermelho representa o valor verdadeiro e as bolinhas verdes representam os valores medidos.

Pode-se verificar que podem existir sistemas de medição com boa acurácia, mas com baixa precisão, caso da Figura 1-b, bem como sistemas com baixa acurácia, mas com boa precisão, caso da Figura 1-c.

No caso da Figura 1-b, embora se saiba que o valor verdadeiro esteja próximo aos valores medidos, mas devido à sua dispersão, existe alto grau de incerteza quanto ao seu valor.

No caso da Figura 1-c, pode parecer que o sistema é bom, pois todas as medições parecem apontar para o mesmo valor, o que pode levar a resultados falsos e enganosos, pois os valores medidos podem estar distantes do valor verdadeiro, conduzindo a decisões de projeto e de engenharia equivocadas.

Sem o conhecimento do valor verdadeiro, os resultados da Figura 1 podem ser apresentados como mostrados na Figura 2.

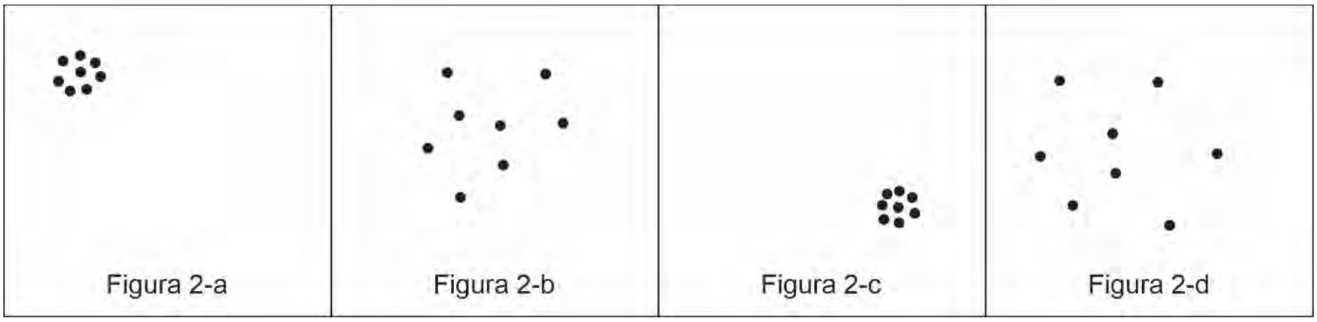



*Figura 2 - Confusão entre acurácia e precisão*
*Fonte: Figura adaptada da Figura 4.9 do documento Survey Methods for Transport Planning*

É fácil de perceber que os sistemas das Figuras 2-a e 2-c apresentam boa precisão. Entretanto, sem ter uma referência do valor verdadeiro, fica difícil saber qual é o sistema com melhor acurácia: o da Figura 2-a ou o da Figura 2-c?

O sistema da Figura 2-b aparenta ter melhor acurácia e precisão do que o da Figura 2-d, mas ainda contém muita incerteza sobre o valor verdadeiro.

Dessa forma, para aferir a qualidade de um sistema de medição, não basta determinar a sua acurácia. É necessário também apurar a sua precisão. Um sistema com baixa precisão torna os resultados obtidos indefinidos em relação ao valor verdadeiro.

Finalmente, deve-se observar que acurácia e precisão são conceitos qualitativos.

O que se quantifica é o erro do sistema de medição para expressar a variação da grandeza medida em relação ao valor verdadeiro. Assim, o erro de um resultado de medição é a diferença do valor medido em relação ao valor verdadeiro. O erro de medição oferece uma ideia da acurácia do sistema.

Outro parâmetro quantificável é o desvio padrão do sistema de medição que reflete a dispersão ou a variabilidade dos valores medidos em relação à sua média. Portanto, o desvio padrão (apresentado no item 3) representa o grau de precisão do sistema.

Parâmetros como o coeficiente de correlação também dão uma ideia do grau de variabilidade dos valores. O conceito de coeficiente de correlação está descrito no item 5.11 do presente trabalho.

# 3. DESVIO PADRÃO POPULACIONAL E AMOSTRAL

O desvio padrão é uma medida de dispersão ou variação dos valores medidos. É uma característica do sistema de medição e é representado por $\boldsymbol{\sigma}$ (**“sigma”** - desvio padrão populacional). Entretanto, o valor verdadeiro do desvio padrão populacional $\boldsymbol{\sigma}$ normalmente não é conhecido (já que a população pode ser muito grande ou infinita). O que se faz é estimar $\boldsymbol{\sigma}$ usando-se uma amostra, obtendo-se, assim, o desvio padrão amostral $\boldsymbol{s}$. Obviamente, quanto maior for a amostra, melhor será a estimativa de $\boldsymbol{s}$ em relação a $\boldsymbol{\sigma}$. Dessa forma, $\boldsymbol{s}$ não é o desvio padrão do sistema de medição, mas é apenas sua estimativa.

O desvio padrão reflete a precisão do sistema de medição. Quanto maior for o desvio padrão, maior é a variabilidade e a dispersão dos dados medidos em relação à sua média e pior será a precisão do sistema.
O desvio padrão amostral $\boldsymbol{s}$ é dado pela expressão:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Onde:

$s$ = desvio padrão da amostra (desvio padrão amostral)
$n$ = número de elementos da amostra
$\bar{x}$ = média aritmética da amostra (média amostral)

## 4. ERRO AMOSTRAL E INTERVALO DE CONFIANÇA

Além dos erros aleatórios e sistemáticos do sistema de medição (que determinam a sua acurácia e precisão), há ainda o erro amostral.

Para medir o erro do sistema de medição de forma exata seria necessária uma amostra infinita. Como isso é impossível, o que se faz é estimar o erro do sistema por meio de uma amostra finita. Assim como não é possível determinar o valor verdadeiro da grandeza que está sendo medida, também não é possível determinar o valor exato do erro do sistema de medição. Novamente, o que se faz é estimar esse erro por meio de uma amostra. O resultado dessa estimativa deve ser o mais próximo possível do verdadeiro valor do erro do sistema de medição. A diferença entre o valor verdadeiro do erro do sistema de medição $E$ e o seu valor estimado $\hat{E}$ é o erro amostral, isto é, $e = E - \hat{E}$, onde $e$ é o erro amostral. Quanto maior for a amostra, menor será o erro amostral e vice-versa.

A Figura 3 ilustra o conceito do erro amostral.

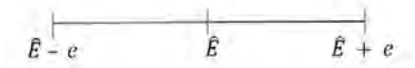



*Figura 3 - Erro amostral e intervalo de confiança*

Na Figura 3, com um nível de confiança de $1 - \alpha$:

- "$\hat{E}$" representa o valor estimado do erro do sistema de medição ($\hat{E}$ é a estimativa de $E$)
- "$e$" representa o erro amostral da estimativa

A chance de o valor verdadeiro $E$ do erro do sistema de medição (desconhecido) estar no intervalo $[\hat{E} - e; \hat{E} + e]$ é de $(1 - \alpha)$ %, isto é, com $(1 - \alpha)$ % de probabilidade tem-se:

$$\hat{E} - e \leq E \leq \hat{E} + e$$

O intervalo $[\hat{E} - e; \hat{E} + e]$ é denominado intervalo de confiança e $(1 - \alpha)$ é o nível de confiança.

O significado de $\alpha$ será detalhado logo adiante.

Quando o erro amostral $e$ for muito grande, a amplitude do intervalo de confiança aumenta, aumentando a incerteza da estimativa.
À primeira vista, pode parecer que o erro amostral não depende do sistema de medição e que não tem relação com a sua acurácia e precisão, dependendo apenas do tamanho da amostra. Entretanto, o erro amostral depende também do desvio padrão do erro do sistema de medição, conforme pode-se ver na Expressão 1.

A expressão que relaciona o erro amostral com o tamanho da amostra e com o nível de confiança é:

$$n = \left(\frac{t_{\alpha/2} \times s}{e}\right)^2 \qquad (1)$$

Onde:

$n$ = tamanho da amostra
$s$ = desvio padrão da amostra
$e$ = erro amostral
$t_{\alpha/2}$ = valor crítico da Distribuição $t$ de *Student* correspondente ao nível de confiança $(1-\alpha)$ desejado e aos graus de liberdade da amostra

Dado um erro amostral e um nível de confiança $(1-\alpha)$, pode-se estabelecer o tamanho da amostra (Expressão 1).

Reciprocamente, dado um tamanho de amostra e um nível de confiança, pode-se estabelecer o erro amostral, conforme a expressão abaixo (derivada da Expressão 1):

$$e = \frac{t_{\alpha/2} \times s}{\sqrt{n}}$$

Por outro lado, o tamanho da amostra depende do desvio padrão amostral $s$. Fixado o erro amostral, quanto maior for o desvio padrão, maior deve ser o tamanho da amostra. Reciprocamente, fixado um tamanho da amostra, quanto maior for o desvio padrão, maior será o erro amostral, aumentando a incerteza quanto ao verdadeiro valor do erro do sistema. Isso significa que sistemas com baixa precisão (maior dispersão dos valores medidos) requerem amostras maiores, sob pena de ter um erro amostral muito grande na estimativa de $E$.

O nível de confiança (1 - $\alpha$) representa a probabilidade de o valor verdadeiro do erro $E$ estar no intervalo $(\hat{E} - e; \hat{E} + e]$, sendo $\hat{E}$ o valor estimado de $E$ (valor verdadeiro do erro do sistema de medição) e $e$ é o erro amostral.

Supondo um nível de confiança de 95%, isto é, quando há 95% de chance de que o valor estimado $\hat{E}$ não divirja do valor verdadeiro $E$ mais do que $\pm e$, tem-se:

$$\alpha = 1 - 0{,}95 = 5\%.$$

Como na Expressão (1) foi adotado o desvio padrão amostral $s$ (em vez do desvio padrão populacional $\sigma$), deve ser adotada a distribuição $t$ de *Student*, em vez da Distribuição Normal.

Para 95% de nível de confiança $(1 - \alpha = 95\%)$, a distribuição $t$ de *Student* fornece a seguinte tabela em função do número de graus de liberdade.

| Graus de liberdade | $t_{5\%/2}$ |
|---|---|
| **20** | 2,0860 |
| **21** | 2,0796 |
| **22** | 2,0739 |
| **23** | 2,0687 |
| **24** | 2,0639 |
| **25** | 2,0595 |
| **26** | 2,0555 |
| **27** | 2,0518 |
| **28** | 2,0484 |
| **29** | 2,0452 |
| **30** | 2,0423 |
| **35** | 2,0301 |
| **40** | 2,0211 |
| **45** | 2,0141 |
| **50** | 2,0086 |
| **60** | 2,0003 |
| **90** | 1,9867 |
| **120** | 1,9799 |
| **240** | 1,9699 |
| **400** | 1,9659 |
| **500** | 1,9647 |

O número de graus de liberdade é $n' - 1$. Por exemplo, para um tamanho de amostra igual a $n' = 24$, tem-se 23 graus de liberdade e, pela tabela, $t_{5\%/2} = 2{,}0687$. Aqui, $n'$ representa uma amostra preliminar, usada especificamente para a determinação do desvio padrão amostral $s$.

O gráfico da Figura 4 ilustra o significado do número 2,0687 do exemplo acima.

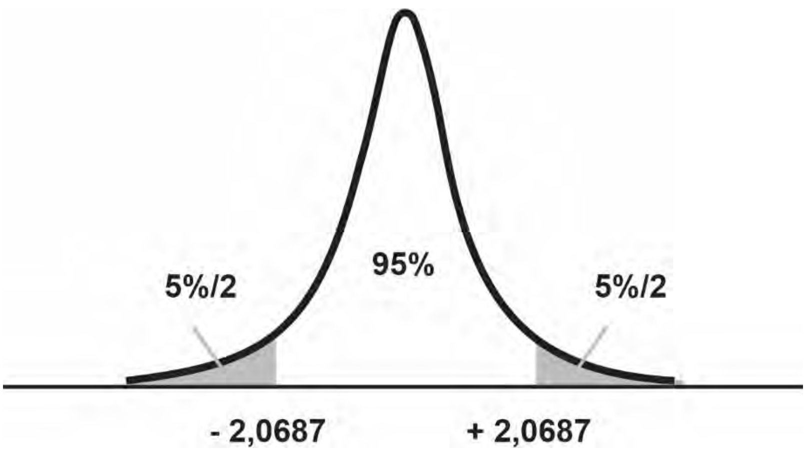



*Figura 4 - Valor de $t_{5\%/2}$ para 23 graus de liberdade*

As duas áreas hachuradas (5%/2 + 5%/2 = 5%) representam a probabilidade de o valor verdadeiro do erro $E$ não estar no intervalo de confiança.

Para amostras grandes (maiores ou iguais a 500) a distribuição $t$ de Student se aproxima da Distribuição Normal, com $t_{5\%/2} \cong Z_{5\%/2} = 1{,}96$, onde $Z_{5\%/2}$ representa o valor crítico para um nível de confiança de 95% na Distribuição Normal.

# 5. METODOLOGIAS PARA A DETERMINAÇÃO DO ERRO DO SISTEMA DE MEDIÇÃO

Como já dito anteriormente, o erro é a diferença entre o valor verdadeiro e o valor medido. Também foi dito que o valor verdadeiro não é conhecido. Então, como se deve calcular o erro na aferição do sistema? Para esse cálculo, usa-se um valor de referência como sendo o valor verdadeiro. Por exemplo, em contagens de veículos motorizados, bicicletas ou pedestres, pode-se usar contagens manuais como valores de referência. Entretanto, deve-se ter em mente que as contagens manuais também podem conter erros, devendo-se ter o cuidado para que o erro da contagem manual não seja significativo o suficiente para comprometer a estimativa do erro do sistema de medição.

Para a determinação do erro de sistemas de medição, não basta calcular o erro de uma única medição. Um único valor não representa o erro do sistema de medição, pois é um valor aleatório e não apresenta nenhum significado real, não havendo vínculo com as características funcionais e operacionais do sistema. Se for feita outra medida, o erro pode ser totalmente distinto. Assim, para a determinação do erro do sistema de medição deve-se fazer um conjunto de medidas. A metodologia é justamente a forma ou o procedimento de como calcular o erro com base nos dados coletados, de forma a refletir com a maior fidelidade possível a real acurácia do sistema de medição.

Em pesquisa na literatura, foram identificadas algumas metodologias para a determinação do erro.

Entre as metodologias ou métricas para a medição de erro encontradas na pesquisa destacam-se as seguintes.

1. *Total Error*
2. *Mean Square Error* - MSE
3. *Root Mean Square Error* - RMSE
4. *Mean Absolute Error* - MAE
5. *Mean Percentage Error* - MPE
6. *Mean Absolute Percentage Error* - MAPE
7. *Weighted Average Percentage Deviation - WAPD*
8. *Symmetric Mean Absolute Percentage Error - sMAPE*
9. *Error per Class - Er*
10. *Mean Error per Class - MEr*
11. ***Pearson's** Correlation Coefficient* - $r$
12. GEH (*Geoffrey E. Havers*)
13. Observação gráfica

Como será observado adiante, algumas metodologias aqui apresentadas utilizam o valor absoluto ou a raiz quadrada do quadrado da diferença. Essas técnicas visam neutralizar as compensações entre as diversas medições, isto é, evitar que erros positivos e negativos se cancelem.

Outra observação importante que deve ser ressaltada é que algumas métricas apresentam unidade (a mesma unidade da grandeza medida ou o seu quadrado), enquanto outras são adimensionais.

As métricas que apresentam unidade não podem ser utilizadas para comparação de medições de grandezas que apresentam unidades diferentes. Por exemplo, não se pode comparar o erro de um contador de veículos com o erro de um contador de pedestres quando a métrica usada para medir o erro apresenta unidade (o erro do contador de veículos é medido em veículos e o erro do contador de pedestres é medido em pedestres). As métricas adimensionais (que não dependem de unidade) podem ser utilizadas para comparar o erro de medições de grandezas com unidades diferentes. Por exemplo, o erro do contador de veículos em % pode ser comparado com o erro do contador de pedestres em %.

As métricas que apresentam unidade são dependentes de escala, enquanto as métricas adimensionais não dependem de escala.

Dependendo da aplicação, é importante a utilização de métricas que sejam independentes de escala. Por exemplo, um erro de 20 veículos numa contagem de 100 veículos é totalmente diferente de um erro de 20 veículos numa contagem de 1000 veículos.

Nas expressões 2 a 12, a seguir, tem-se:

$x_i$ = valor medido na $i$-ésima medição

$y_i$ = valor "verdadeiro" (valor de referência) da $i$-ésima medição

$n$ = número de amostras

## 5.1 *Total Error*

O "*total error*" é a diferença entre a soma dos valores medidos e a soma dos valores verdadeiros ou de referência, dividido pela soma dos valores verdadeiros e é calculado conforme a Expressão 2, sendo uma medida adimensional e independente de escala.

$$Total\ Error = \frac{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i} \qquad (2)$$

O *Total Error* pode ser negativo, indicando uma subcontagem (no caso de sistemas de contagem).

É uma métrica que faz a compensação entre erros positivos e negativos, de forma que sobrecontagens e subcontagens podem se anular mutuamente.

## 5.2 *Mean Square Error* - MSE

O MSE é a média dos quadrados dos erros.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 \qquad (3)$$

É uma métrica que apresenta unidade (a mesma unidade da grandeza medida ao quadrado) e é dependente de escala. Por exemplo, se $x_i$ e $y_i$ forem medidos em veículo, a unidade do MSE será veículo$^2$ (veículo ao quadrado).

É uma métrica que não faz a compensação entre erros positivos e negativos. Sobrecontagens e subcontagens não se anulam mutuamente.

## 5.3 *Root Mean Square Error* - RMSE

O RMSE é também conhecido como NRMSE - *Normalized Root Mean Square Error*.

O RMSE é a raiz quadrada do MSE:

$$RMSE = \sqrt{MSE}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (4)$$

É uma métrica que apresenta unidade (a mesma unidade da grandeza medida) e é dependente de escala. Se $x_i$ e $y_i$ forem medidos em veículos, no caso do MSE, a unidade será veículo$^2$ (veículo ao quadrado) e no caso do RMSE a unidade será veículo (a mesma unidade da grandeza medida).

É uma métrica que não faz a compensação entre erros positivos e negativos.

## 5.4 *Mean Absolute Error* - MAE

O MAE é a média dos valores absolutos dos erros. É também conhecido como *Mean Absolute Deviation* - MAD e é calculado conforme a Expressão 5.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \qquad (5)$$

É uma métrica que apresenta unidade (a mesma unidade da grandeza medida) e é dependente de escala.

É uma métrica que não faz a compensação entre erros positivos e negativos.

## 5.5 *Mean Percentage Error* - MPE

O MPE também é denominado de *Average Percentage Deviation* - APD.

O MPE representa o erro médio em termos percentuais e é calculado conforme a Expressão 6.

$$MPE = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - y_i)}{y_i} \qquad (6)$$

Essa métrica faz compensação entre erros positivos e negativos. Sobrecontagens e subcontagens podem se anular mutuamente.

O MPE pode resultar em valor negativo, indicando uma subcontagem.

É uma grandeza adimensional e não depende de escala.

A desvantagem desse método é que erros em situações de baixo volume representam altas porcentagens, enviesando o resultado. Por exemplo, uma sobrecontagem de 1 em relação a um volume de 1 representa um erro de 100%. No limite, quando o valor verdadeiro $y_i$ for igual a zero, a métrica não fica aplicável (pois não se pode dividir por zero).

Outra desvantagem é que o MPE não é simétrico: sobrecontagens têm peso maior do que subcontagens. Por exemplo, uma subcontagem de 1 em um volume de 2 (erro de -1) representa um erro de - 50%, enquanto uma sobrecontagem de 1 em um volume de 1 (erro de +1) resulta em um MPE de +100%. Assim, um sistema que faz subcontagens terá a tendência de apresentar um MPE em valor absoluto menor do que um sistema que faz sobrecontagens.

## 5.6 *Mean Absolute Percentage Error* - MAPE

O MAPE também é denominado por *Absolute Interval Error* - AIE ou, ainda, por *Average of the Absolute Percentage Deviation* - AAPD.

O MAPE é a média do valor absoluto dos erros em termos percentuais, contornando o problema das compensações entre erros positivos e negativos apresentado pelo MPE e é calculado pela Expressão 7.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - y_i}{y_i}\right| \qquad (7)$$

O MAPE é adimensional e não depende de escala.

O MAPE tem um limite inferior de 0%, mas não apresenta limite superior.

A desvantagem desse método é que erros em situações de baixo volume representam altas porcentagens, enviesando o resultado. Por exemplo, um erro de 1 em relação a um volume de 1 representa um erro percentual de 100%, enquanto um erro de 1 em um volume de 100 representa um erro percentual de 1%. No limite, quando o valor verdadeiro $y_i$ for igual a zero, a métrica não é aplicável (pois não se pode dividir por zero).

O MAPE é assimétrico e penaliza erros em que o valor medido é maior do que o valor verdadeiro. Isso é devido ao fato de que o erro percentual não pode exceder 100% no caso em que o valor medido é menor que o valor verdadeiro[1]. Por outro lado, quando o valor medido é maior do que o valor verdadeiro, não há limite superior para o erro percentual. Como resultado, MAPE favorece subcontagens e penaliza sobrecontagens. Isto é, um sistema

[1] O MAPE atinge 100% quando $y_i = 2x_i, \forall i$.

que faz subcontagens poderá ter um MAPE melhor do que outro que faz sobrecontagens. Veja no quadro abaixo um exemplo numérico:

| | $x_i$ (valor medido) | $y_i$ (valor verdadeiro) | Erro absoluto | MAPE |
|---|---|---|---|---|
| Sistema A | 100 | 20 | 80 | 400% |
| Sistema B | 20 | 100 | 80 | 80% |

Pelo exemplo numérico acima, verifica-se, para um mesmo erro absoluto, o MAPE penalizou o sistema com sobrecontagem em relação ao sistema com subcontagem. No Sistema A houve um erro de 80 em 20 (400%), enquanto no Sistema B ocorreu um erro de 80 em 100 (80%). Assim, o erro relativo é maior em sobrecontagens.

Entretanto, se o valor verdadeiro for igual tanto para subcontagem como para sobrecontagem, o MAPE permanece igual, como mostrado no exemplo numérico abaixo:

| | $x_i$ (valor medido) | $y_i$ (valor verdadeiro) | Erro absoluto | MAPE |
|---|---|---|---|---|
| Sistema A | 180 | 100 | 80 | 80% |
| Sistema B | 20 | 100 | 80 | 80% |

## 5.7 *Weighted Average Percentage Deviation - WAPD*

O WAPD também é conhecido como *Weighted Absolute Interval Error* - WAIE, *Weighted Mean Absolute Percentage Error* - WMAPE ou ainda *Weighted Absolute Percentage Error* - WAPE e é calculado conforme a Expressão 8.

$$WAPD = \sum_{i=1}^{n} \frac{|x_i - y_i|}{\sum_{j=1}^{n} y_j} = \frac{\sum_{i=1}^{n} |x_i - y_i|}{\sum_{i=1}^{n} y_i} \qquad (8)$$

É uma métrica que não faz a compensação entre erros positivos e negativos.

O WAPD é adimensional e não depende de escala.

O WAPD atenua o problema de altos erros percentuais em situações de baixo volume existente no MAPE, conforme mostra o exemplo numérico abaixo.

| **Intervalo $i$** | $x_i$ | $y_i$ | $\left\lvert\frac{x_i - y_i}{y_i}\right\rvert$ | $\frac{\lvert x_i - y_i\rvert}{\sum_{i=1}^{n} y_i}$ |
|---|---|---|---|---|
| **1** | 11 | 5 | 120,00% | 15,38% |
| **2** | 10 | 8 | 25,00% | 5,13% |
| **3** | 17 | 15 | 13,33% | 5,13% |
| **4** | 12 | 11 | 9,09% | 2,56% |
| | $\sum x_i = 50$ | $\sum y_i = 39$ | MAPE = 52,78% | WAPD = 28,21% |

No exemplo numérico, o WAPD resultou em um valor menor que o MAPE.

Ao contrário do MAPE, o WAPD pode ser aplicado mesmo quando houver valores de $y_i$ iguais a zero.

Observação 1:

A Expressão 8 é derivada da seguinte expressão:

$$WAPD = \sum_{i=1}^{n} \left( \left| \frac{x_i - y_i}{y_i} \right| \times \frac{y_i}{\sum_{j=1}^{n} y_j} \right)$$

Observação 2:

No artigo "*Performance Evaluation and Correction Functions for Automated Pedestrian and Bicycle Counting Technologies*", o *Weighted Average Percentage Deviation - WAPD* é designado como *Weighted Average Absolute Percentage Deviation* - WAAPD. Por outro lado, esse artigo designa como WAPD - *Weighted Average Percentage Deviation* a seguinte expressão:

$$WAPD = \sum_{i=1}^{n}\left(\frac{x_i - y_i}{y_i} \times \frac{y_i}{\sum_{j=1}^{n} y_j}\right)$$

## 5.8 *Symmetric Mean Absolute Percentage Error - sMAPE*

O sMAPE é calculado pela Expressão 9:

$$sMAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - y_i|}{\frac{(x_i + y_i)}{2}} \qquad (9)$$

O sMAPE também é expresso em porcentagem, não apresenta unidade e é independente de escala.

O sMAPE possui um limite inferior de 0% e um limite superior de 200%. Sempre que o valor medido ou o valor verdadeiro for igual a zero, o sMAPE atingirá 200%.

O sMAPE contorna o problema de assimetria do MAPE, conforme mostra o exemplo numérico abaixo.

| | $x_i$ (valor medido) | $y_i$ (valor verdadeiro) | Erro absoluto | MAPE | sMAPE |
|---|---|---|---|---|---|
| **Sistema A** | 100 | 20 | 80 | 400% | 133,3% |
| **Sistema B** | 20 | 100 | 80 | 80% | 133,3% |

Ao mesmo tempo que o sMAPE atenua o problema de assimetria do MAPE, é introduzido outro tipo de assimetria causada pelo denominador da Expressão 9. Veja o exemplo numérico abaixo:

| $x_i$ (valor medido) | $y_i$ (valor verdadeiro) | Erro absoluto | MAPE | sMAPE |
|---|---|---|---|---|
| **120** | 100 | 20 | 20% | 18,2% |
| **80** | 100 | 20 | 20% | 22,2% |

No exemplo numérico acima, verifica-se que, para o mesmo erro absoluto e para um mesmo valor verdadeiro, o sMAPE resultou em valores distintos, penalizando os casos em que a soma (valor medido + valor verdadeiro) é menor.

A faixa de 0% a 200% não é fácil de interpretar, pois não é intuitiva. Por isso, muitas vezes o fator 2 é omitido da Expressão 9 (para que a faixa de valores fique em 0% a 100%). Neste caso, o sMAPE não será comparável com o MAPE.

Embora o sMAPE não apresente o problema do MAPE quando o valor verdadeiro é zero, apresenta o mesmo problema quando ambos os valores (medido e verdadeiro) são iguais a zero, apesar de que pode se convencionar que, neste caso, o erro percentual seja 0%. Entretanto, o sMAPE fica instável quando ambos os valores (medido e verdadeiro) são próximos de zero (ou quando um deles for zero e o outro próximo de zero), resultando em porcentagens elevadas para erros pequenos.

## 5.9 *Error per Class* - *Er* e *Mean Error per Class* - $MEr$

### 5.9.1 *Error per Class* - *Er*

Normalmente, em contagens classificadas, o volume de veículos de uma determinada classe pode ser muito baixo, tão baixo que mesmo o WAPD pode apontar erros percentuais significativamente grandes para erros pequenos.

O $Er$ representa o erro de classificação nas seguintes situações:

a) Detecção incorreta quando o sistema não registra um veículo;
b) Falsa detecção quando o sistema registra um veículo inexistente; ou
c) Combinação de detecção incorreta e detecção falsa.

O $Er$ é calculado pela Expressão 10.

$$Er = 1 - \frac{\min(X;Y)}{\max(X;Y)} \qquad (10)$$

Onde:

$X = \sum_{i=1}^{n} x_i$

$Y = \sum_{i=1}^{n} y_i$

*Fonte: The Study of Vehicle Classification Equipment with Solutions to Improve Accuracy in Oklahoma, Final Report - FHWA-OK-14-17 (2014).*

O $Er$ resolve o problema de altos erros percentuais em situações de baixíssimo volume que mesmo o WAPD não consegue resolver, conforme ilustrado no exemplo numérico abaixo:

| $i$ | $x_i$ | $y_i$ | $\frac{\lvert x_i - y_i \rvert}{\sum_{i=1}^{n} y_i}$ | |
|---|---|---|---|---|
| 1 | 1 | 2 | 20,00% | |
| 2 | 3 | 1 | 40,00% | |
| 3 | 2 | 1 | 20,00% | |
| 4 | 0 | 0 | 0,00% | |
| 5 | 1 | 1 | 0,00% | |
| Total | 7 | 5 | WAPD = 80,00% | *Er* = 28,57% |

### 5.9.2 *Mean Error per Class* - *MEr*

Com base em $Er$ pode-se construir um indicador médio, aqui denominado de $MEr$, calculado conforme a Expressão 11:

$$MEr = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\min(x_i, y_i)}{\max(x_i, y_i)}\right) \qquad (11)$$

O valor médio $MEr$ representa com maior fidelidade o comportamento do sistema de medição do que o $Er$ (que considera apenas os valores totais), pois incorpora as informações contidas nos intervalos, conforme o exemplo numérico abaixo.

| Sistema A | | | | | Sistema B | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $i$ | $x_i$ | $y_i$ | $1 - \frac{\min(x_i,\, y_i)}{\max(x_i,\, y_i)}$ | | $i$ | $x_i$ | $y_i$ | $1 - \frac{\min(x_i,\, y_i)}{\max(x_i,\, y_i)}$ | |
| 1 | 11 | 5 | 54,55% | | 1 | 7 | 6 | 14,29% | |
| 2 | 10 | 8 | 20,00% | | 2 | 10 | 10 | 0,00% | |
| 3 | 17 | 15 | 11,76% | | 3 | 14 | 13 | 7,14% | |
| 4 | 12 | 20 | 40,00% | | 4 | 19 | 19 | 0,00% | |
| | $\sum x_i = 50$ | $\sum y_i = 48$ | $MEr$ = 31,58% | $Er$ = 4,00% | | $\sum x_i = 50$ | $\sum y_i = 48$ | $MEr$ = 5,36% | $Er$ = 4,00% |

No exemplo, como os valores totais foram iguais (50 e 48), o $Er$ foi igual para os dois sistemas (4,00%). Entretanto, pode-se perceber que a acurácia do Sistema B é melhor do que a do Sistema A. Assim, o $MEr$ retratou melhor a acurácia do que o $Er$.

Em sistemas que apresentam erros sistemáticos de subcontagem, o $MEr$ será próximo do MAPE. No limite, quando houver subcontagem em todos os intervalos, as duas métricas apresentarão o mesmo resultado, conforme o exemplo numérico abaixo.

| Intervalo | $x_i$ | $y_i$ | Erro absoluto | $\left\lvert\frac{x_i - y_i}{y_i}\right\rvert$ | $1 - \frac{\min(x_i,\, y_i)}{\max(x_i,\, y_i)}$ |
|---|---|---|---|---|---|
| 1 | 34 | 39 | 5 | 12,82% | 12,82% |
| 2 | 30 | 34 | 4 | 11,76% | 11,76% |
| 3 | 29 | 32 | 3 | 9,38% | 9,38% |
| 4 | 31 | 33 | 2 | 6,06% | 6,06% |
| 5 | 28 | 29 | 1 | 3,45% | 3,45% |
| | | | | MAPE = 8,69% | $MEr$ = 8,69% |

Segue abaixo um quadro resumo com as principais características das métricas apresentadas.

| Métricas | Adimensional | Dependente de escala | Positivo / Negativo | Compensação de erros positivos e negativos | Simetria | Alto erro percentual com baixo volume |
|---|---|---|---|---|---|---|
| *Total Error* | Sim | Não | P / N | Sim | - | Não |
| *MSE* | Não | Sim | P | Não | Sim | Não |
| *RMSE* | Não | Sim | P | Não | Sim | Não |
| *MAE* | Não | Sim | P | Não | Sim | Não |
| *MPE* | Sim | Não | P / N | Sim | Não | Sim |
| *MAPE* | Sim | Não | P | Não | Não | Sim |
| *WAPD* | Sim | Não | P | Não | Sim | Não* |
| *sMAPE* | Sim | Não | P | Não | Sim | Não |
| *Er* | Sim | Não | P | Sim | Sim | Não |
| *MEr* | Sim | Não | P | Sim | Sim | Não* |

* Sim, para volumes extremamente baixos (como em contagens classificadas).

### 5.10 *Pearson's Correlation Coefficient* ($r$)

> *"O que a correlação procura entender é como uma variável se comporta em um cenário onde outra está variando, visando identificar se existe alguma relação entre a variabilidade de ambas. Embora não implique em causalidade, o coeficiente de correlação exprime em números essa relação, ou seja, quantifica a relação entre as variáveis.*
>
> *Mas não existe apenas uma forma de se calcular a correlação entre variáveis. Dependendo da forma e de como se comportam as variáveis, um coeficiente de correlação é mais adequado que outro.*
>
> *O coeficiente de correlação de Pearson, também chamado de correlação linear ou $r$ de Pearson, é um grau de relação entre duas variáveis quantitativas e exprime o grau de correlação através de valores situados entre - **1** e **1**."*
>
> *Fonte: Coeficientes de correlação: Para que servem e como interpreta-los? (operdata.com.br)*

O coeficiente de correlação de *Pearson* indica quão correlacionadas entre si estão duas variáveis. $r = \pm 1$ indica uma perfeita correlação entre as duas variáveis, enquanto $r = 0$ indica que não há nenhuma correlação entre as duas variáveis.

No presente trabalho é proposto o uso do coeficiente de *Pearson* $r$ para exprimir o grau de precisão do sistema de medição. Quanto mais próximo $r$ estiver de +1, melhor será a precisão do sistema.

Foi dito anteriormente que o desvio padrão reflete a dispersão ou a variabilidade dos valores e que, portanto, representa o grau de precisão do sistema. Então, por que não se adota o desvio padrão em vez do coeficiente de correlação de *Pearson* para exprimir o grau de precisão do sistema?

A resposta é que o desvio padrão não possui um limite superior e não apresenta um valor que possa servir como referência. Por falta de um limite superior, fica difícil de interpretar se um resultado é satisfatório ou não, assim como, por falta de uma referência, fica difícil de estabelecer valores limites para aprovação ou rejeição. O coeficiente de correlação de Pearson $r$ apresenta um limite superior (+1) e quanto $r$ estiver mais próximo de +1, melhor será a precisão. Assim, pode-se convencionar valores limites para aprovação ou rejeição, de acordo com a precisão necessária de cada aplicação.

O coeficiente de *Pearson* $r$ é calculado conforme a Expressão 12.

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{Y})(x_i - \bar{X})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{Y})^2}\sqrt{\sum_{i=1}^{n}(x_i - \bar{X})^2}} \qquad (12)$$

Onde:
$\bar{X}$ = média aritmética dos valores medidos
$\bar{Y}$ = média aritmética dos valores verdadeiros ou valores de referência
O Excel calcula automaticamente o coeficiente de correlação de *Pearson*, com a função correl (matriz 1; matriz 2) ou *Pearson*(matriz 1; matriz 2).

5.11 GEH (*Geoffrey E. Havers*)

$$GEH = \sqrt{\frac{(M-C)^2}{\frac{M+C}{2}}} \qquad (13)$$

Onde:

$M$ = valor medido
$C$ = valor verdadeiro (valor de referência)
A condição de aderência aceitável dos valores medidos com os valores verdadeiros é:

$$GEH < 5$$

Note que GEH possui unidade. Segundo o artigo "*The GEH Measure and Quality of the Highway Assignment Models*", para haver aderência entre o valor medido e o valor verdadeiro, a condição $GEH < 5$ deve ser válida em 85% dos casos e tanto $M$ como $C$ devem estar expressos em veículos/hora.

Segundo *Traffic Modelling Guidelines* - TfL 2010: "A estatística GEH é uma medida padrão da "qualidade do ajuste" entre fluxos observados e modelados. Ao contrário da comparação de fluxos usando a diferença percentual, a estatística GEH coloca mais ênfase em fluxos maiores do que em fluxos menores."

5.11 Observação gráfica e análise qualitativa

Para o método de observação gráfica, as duas contagens são plotadas em cima de uma linha de 45° para fazer uma avaliação visual e qualitativa da variação dos dados, conforme mostra a Figura 3. Visualmente, o resultado da Figura 3 indica que há uma boa aderência entre os dados, já que os pontos plotados situam-se próximos à reta de 45°.

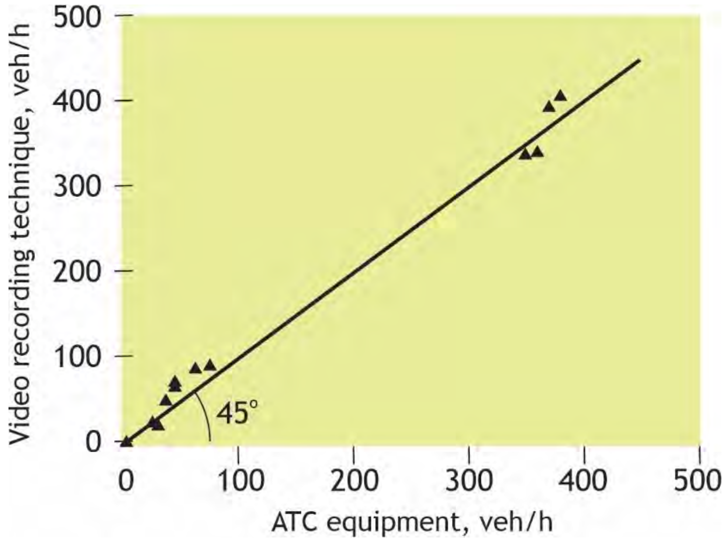



*Figura 3 - Observação gráfica*
*Fonte: Applicability of an Automatic Pneumatic-Tube-Based Traffic Counting Device for Collecting Data Under Mixed Traffic*

# 6. SUGESTÕES E RECOMENDAÇÕES

Para sistemas de medição que fazem a contagem de veículos, pedestres ou bicicletas recomenda-se o uso das seguintes metodologias ou métricas, as quais devem ser consideradas em conjunto. Isso porque, conforme foi visto no item 5 do presente trabalho, cada métrica possui vantagens e desvantagens, sendo que nenhuma métrica, sozinha, consegue retratar com fidelidade toda a gama de situações possíveis de serem encontradas no mundo real.
Além disso, recomendou-se aqui o uso apenas de métricas adimensionais e independentes de escala e que não façam compensações entre erros positivos e erros negativos.

## 6.1 Metodologias de avaliação da acurácia

### 6.1.1 MAPE - *Mean Absolute Percentage Error* (ver a Expressão 7 no item 5.6 deste trabalho).

Essa métrica é recomendada principalmente por duas razões:
a) O MAPE é de fácil utilização e aplicação.
b) O entendimento e a interpretação do resultado do MAPE são simples e intuitivos (é a média do valor absoluto dos erros individuais em termos percentuais).

Por conta dessas vantagens, o MAPE é amplamente encontrado na literatura e utilizado em artigos e em pesquisas.
Entretanto, o MAPE também apresenta importantes desvantagens, sendo que a principal delas ocorre em situações em que o valor verdadeiro é muito baixo. O MAPE é uma média de porcentagens. Se o valor verdadeiro ou de referência for muito baixo, qualquer erro representará uma porcentagem muito grande. Por exemplo, se o valor verdadeiro for 2 e o valor medido for 3, o erro percentual é de 1/2 = 50%. No limite, se o valor verdadeiro for zero, a porcentagem será infinita, já que não se pode dividir por zero. Dessa forma, no procedimento de avaliação, deve-se prever mecanismos para contornar essa situação sem criar viés na estimativa.
Outra desvantagem do MAPE é a assimetria, apresentando erros percentuais diferentes para sobrecontagens e subcontagens, conforme já exposto no item 5.6 desse trabalho.

6.1.2 WAPD - *Weighted Average Percentage Deviation*

Para medições de fluxos mais baixos, recomenda-se aqui usar o WAPD (Expressão 8 do item 5.7 do presente trabalho).

6.1.3 sMAPE - *Symmetric Mean Absolute Percentage Error*

Para evitar casos que penalizem erros de sobrecontagem, recomenda-se usar o sMAPE (Expressão 9 do item 5.8 do presente trabalho).

6.1.4 $\boldsymbol{Er}$ e $\boldsymbol{MEr}$ - *Error per Class* e *Mean Error per Class*

Para contagens muito baixas, como aquelas que que podem ser obtidas em contagens classificadas, recomenda-se o uso de $Er$ e $MEr$ (Expressões 10 e 11 dos itens 5.9.1 e 5.9.2, respectivamente).

6.1.5 Casos em que ambos os valores, medido e verdadeiro, são nulos

Foi visto no item 6.1.1 que pode haver situações de divisão por zero no caso do MAPE, quando o valor verdadeiro for nulo e o valor medido for diferente de zero, isto é, $m/0$, com $m \neq 0$. Neste caso, não é possível calcular o MAPE[2].

Isso não ocorre com as métricas sMAPE e $MEr$. Entretanto, se ambos os valores de um intervalo, medido e verdadeiro, forem nulos, o denominador desses indicadores será igual a zero e ter-se-á uma situação do tipo 0/0. Como o resultado de 0/0 é indefinido[3], isto é, o resultado dessa divisão pode ser qualquer número, pode-se convencionar que, neste caso, o sMAPE e $MEr$ são nulos (uma vez que o erro absoluto é nulo) no intervalo considerado.

## 6.2 Metodologia para a avaliação da precisão

6.2.1 Coeficiente de correlação de *Pearson*

Para a avaliação da precisão, é recomendada aqui a utilização do coeficiente de correlação de *Pearson* (Expressão 12 do item 5.10 deste trabalho).

## 6.3 Erro amostral e tamanho da amostra

Quanto ao erro amostral e tamanho da amostra, recomenda-se, sempre que possível, que se fixe um erro amostral máximo admissível para então determinar a amostra necessária. Para o cálculo do tamanho da amostra é necessário antes determinar o desvio padrão amostral $s$, a partir de uma amostra preliminar.

## 6.4 Procedimento de avaliação

A seguir, recomenda-se o seguinte procedimento de avaliação.

6.4.1 Métricas e critérios para a verificação da acurácia e precisão

As metodologias e critérios para a verificação da acurácia e precisão apresentados a seguir são meramente sugestões dos autores da presente Nota Técnica, podendo ser consideradas outras metodologias ou critérios.

Acurácia

Sugere-se como critério adotar o menor valor obtido para as métricas mencionadas no item 6.1, isto é:

$$\mathrm{A} = \min\,(MAPE; WAPD; sMAPE; Er; MEr) \qquad (14)$$

O critério sugerido é:

2 Pela definição de divisão: se $a/b = c$ ($a \neq 0$ e $c \neq 0$), então $b \times c = a$. Se $b = 0$, não existe $c$ tal que $0 \times c = a$.

3 Pela definição de divisão: se $a/b = c$ ($a = 0$), então $b \times c = 0$. Se $b = 0$, então $0 \times c = 0$, $\forall c$.

$$\begin{cases} A \leq A_1 \rightarrow Sistema\ aprovado\ quanto\ \text{à}\ acurácia \\ A_2 \geq A > A_1 \rightarrow Repetir\ o\ teste \\ A > A_2 \rightarrow Sistema\ reprovado \end{cases}$$

Os valores de $A_1$ e $A_2$ deverão ser previamente definidos.

Precisão
Sugere-se adotar o coeficiente de correlação de *Pearson* $r$ para avaliar a precisão do sistema. Conforme já descrito no item 5.10, o coeficiente de *Pearson* fornece uma medida da relação entre a variabilidade dos valores medidos em relação à variabilidade dos valores verdadeiros ou de referência.
O critério sugerido é:

$$\begin{cases} 1 \geq r \geq r_1 \rightarrow Sistema\ aprovado\ quanto\ \text{à}\ precisão \\ r_2 \leq r < r_1 \rightarrow Repetir\ o\ teste \\ r < r_2 \rightarrow Sistema\ reprovado \end{cases}$$

Os valores de $r_1$ e $r_2$ deverão ser previamente definidos.

Poderão ser adotados valores distintos de $A_1$, $A_2$, $r_1$ e $r_2$ para aplicações distintas. Por exemplo, um conjunto ($A_1$, $A_2$, $r_1$ e $r_2$) de valores para a contagem de pedestres, um conjunto de valores para contagem de bicicletas, um conjunto de valores para cada classe de contagem classificada, ou um conjunto para uma combinação de aplicações.

6.4.2 Repetição de teste
Após a repetição do teste, o critério de avaliação será:

- $A' \leq A_1$
- $1 \geq r' \geq r_1$

Onde:
$A'$ = Resultado da Expressão 14 do segundo teste
$r'$ = Coeficiente de correlação de Pearson do segundo teste

6.4.3 Amostra

I. Usar intervalos de controle de 15 minutos para contagens volumétricas.

II. Para calcular o tamanho da amostra ou o erro amostral sugere-se basear no menor valor das seguintes métricas, relativas à média dos erros absolutos: $MAPE; sMAPE; MEr$, isto é:

$$\min\ (MAPE; sMAPE; MEr)$$

O $Er$ e o *WAPD* não são médias dos erros e, portanto, não foram considerados para efeitos de erro amostral.

III. Calcula-se o valor do tamanho da amostra necessário $n$ pela Expressão 1, usando-se uma amostra preliminar $n'$ para a determinação do desvio padrão amostral $s$ da métrica que apresentou o menor valor, conforme indicado no item II acima. Com base nesse desvio padrão amostral, calcular $n$ usando-se a Expressão 1 (adotar nível de confiança de 95% e um erro amostral máximo admissível de 2%).

a) Se $n \leq n'$, então todas as métricas deverão ser obtidas da própria pesquisa preliminar.
b) Se $n > n'$, então deve-se complementar a pesquisa até atingir-se os $n$ intervalos.

IV. Se $n$ resultar em um valor muito grande, pelo fato de o desvio padrão amostral $s$ ser muito grande, recomenda-se fazer uma análise para verificar a causa de haver desvios tão grandes.

O equipamento pode contar quando não deveria ter contado e não contar quando deveria ter contado.

Isso pode ser devido a:

a) Deficiências, falhas ou erros de calibração/ajustes do sistema de contagem. Neste caso, deverá ser verificado se a deficiência ou a falha identificada pode ou não ser corrigida. Em caso afirmativo, deverá ser refeita a pesquisa.
b) Problemas do local de contagem que conduzem aos erros verificados, que não podem ser caracterizados como erros do sistema de contagem. Esses problemas podem ser temporários (que ocorreram em um ou outro intervalo) ou podem ser predominantes ou contínuos. Neste caso, para:
   - Problema temporário: excluir os intervalos em que o problema ocorreu.
   - Problema predominante ou contínuo: refazer a pesquisa deslocando o ponto de contagem um pouco para frente ou para trás, de forma a evitar o problema verificado.

No caso de não ter sido identificado nenhum problema ou não ser possível a adoção de nenhuma solução conforme apontado no item IV acima, usar uma amostra mínima de 48 intervalos de 15 minutos ($n = 48$). Neste caso, calcular o erro amostral para essa amostra de 48 intervalos para a métrica que tenha apresentado o menor valor, conforme indicado no item II.

V. Outro problema que pode ocorrer é quando o valor verdadeiro ou o valor de referência for muito baixo.

Se o valor baixo ocorrer apenas em um ou outro intervalo, o mesmo pode ser descartado da amostra.

### 6.4.4 Análise gráfica

Recomenda-se sempre fazer uma análise gráfica dos dados, por meio de dois gráficos:

(i) Um gráfico mostrando o comportamento dos valores medidos e dos valores de referência ao longo do tempo (série temporal); e
(ii) Um gráfico de dispersão dos dados.

Exemplos desses gráficos estão mostrados no item 7 adiante.

### 6.4.5 Inteligência do sistema de contagem

No caso específico de contagem de pedestres, deve ser verificado se o sistema tem a habilidade de distinguir os pedestres de animais, objetos (como postes e mobiliário urbano), carrinhos de bebê, entre outros. Outras situações a observar quanto ao desempenho do sistema é se ele consegue contar pedestres com guarda-chuvas, situações com pelotão de pedestres etc.

No caso de sistemas que usam imagens, verificar o desempenho do sistema em diversas situações ambientais e de iluminação: dia/noite; farol de veículo, incidência direta de luz solar, chuva, neblina, sombra etc.

### 6.4.6 Situações de alto fluxo

Deve ser verificado o desempenho do sistema em situações com alto fluxo. No caso de contagem de veículos, é recomendável verificar o desempenho do sistema em situações de lentidão e de fila.

6.4.7 Configuração, calibração e ajustes

Recomenda-se que, durante o processo de avaliação, seja verificado com cuidado o procedimento de configuração, calibração e ajuste do sistema. Na medida do possível, deve-se evitar sistemas com mecanismos complexos e trabalhosos de configuração, calibração e ajustes.

Um sistema com mecanismo complexo de configuração, calibração e ajustes pode passar no processo de avaliação com boa acurácia e alta precisão. Mas, no regime normal de operação, pode haver falhas de configuração, calibração e ajustes (devido à sua complexidade e/ou sensibilidade), gerando erros sistemáticos.

Essa situação pode provocar um resultado igual ao da Figura 1-c, com boa precisão e nenhuma acurácia. Por falta de referência (pois não haverá contagens manuais), o falso resultado será aceito como verdadeiro, levando a decisões equivocadas de planejamento, de projeto ou de engenharia.

# 7 EXEMPLOS DE APLICAÇÃO

Embora a recomendação seja a realização de testes com uma amostra mínima de 48 intervalos de 15 minutos, apenas a título de exemplo, neste item fazem-se exercícios de aplicação utilizando-se exemplos simulados, com uma pesquisa preliminar de $n' = 24$ intervalos de 15 minutos.

Nos 6 exemplos numéricos apresentados neste item 7, serão adotados os seguintes valores, apenas para efeitos de exemplo e exclusivamente a título de ilustração:

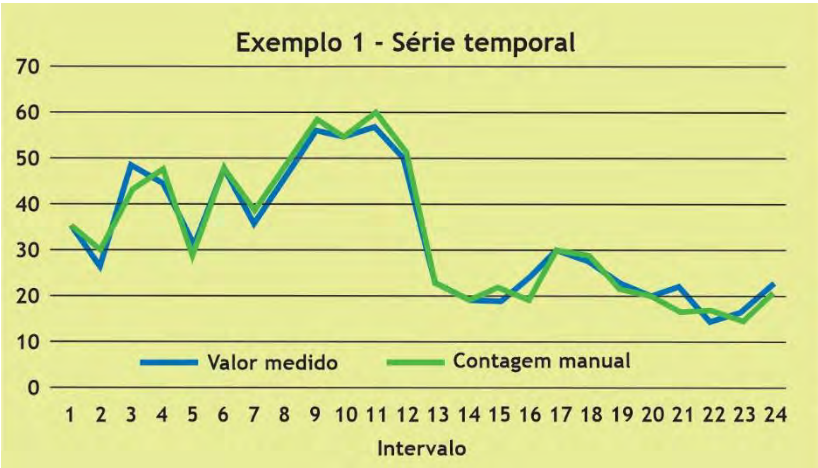
$A_1 = 17\%$

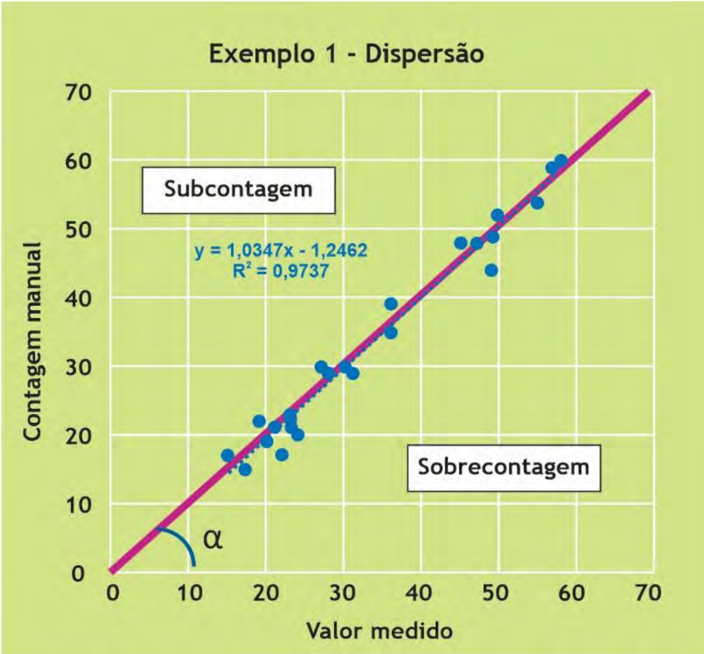
$A_2 = 20\%$

$r_1 = 90\%$

$r_2 = 80\%$

## 7.1 Exemplo 1






No gráfico da dispersão, a reta lilás é uma reta de 45° ($\alpha = 45°$). No eixo horizontal estão representados os valores medidos e no eixo vertical os valores da contagem manual (considerados como valores verdadeiros), isto é, cada ponto azul é representado por um par $(x, y)$, onde $x$ é o valor medido e $y$ é o valor da contagem manual. A reta lilás é a reta de todos os pontos onde $x = y$. Se todos os valores medidos fossem iguais aos valores da contagem manual, todos os pontos azuis estariam na reta lilás e a reta azul pontilhada coincidiria com a reta lilás. Fazendo analogia com tiro ao alvo, é como se todos os tiros acertassem o alvo em cheio.

A reta lilás do gráfico de dispersão representa o centro do alvo.

Se os pontos azuis ficarem distantes da reta lilás, é como se os tiros ficassem longe do centro do alvo.

A reta azul pontilhada do gráfico de dispersão é a reta de regressão linear. No exemplo, a equação dessa reta é $y = 1{,}0347x - 1{,}2462$ (calculada pelo Excel). O coeficiente $R^2$ é o coeficiente de determinação e é uma medida que exprime quão próximos os dados estão da linha de regressão ajustada (a reta azul pontilhada). $R^2$ é um valor compreendido entre 0 e 1. Quanto mais próximo estiver de 1, mais os pontos vão estar próximos da reta azul pontilhada (não necessariamente da reta lilás). No exemplo, tem-se $R^2 = 0{,}9737$, mostrando que os pontos estão bem próximos da reta azul pontilhada. O $R^2$ é o quadrado do coeficiente de correlação de *Pearson* $r$.

Note que um $R^2$ próximo de 1 não significa necessariamente boa acurácia, representando apenas boa precisão. De fato, é possível uma situação em que todos os pontos azuis estejam em cima da reta azul pontilhada ($R^2$ = 1), mas a reta azul pontilhada esteja longe da reta lilás, estando, portanto, os valores medidos muito afastados dos valores verdadeiros, resultando em uma baixa acurácia.

Pela análise visual dos gráficos, verifica-se que houve boa aderência dos valores medidos com os valores da contagem manual.

Quanto à análise dos indicadores:

| MAPE | WAPD | sMAPE | $MEr$ (médio) | $Er$ (total) | Min | | Correlação $r$ | | Análise gráfica | Avaliação final |
|---|---|---|---|---|---|---|---|---|---|---|
| **7,10%** | 5,73% | 6,90% | 6,50% | 0,25% | 0,25% | S | 98,67% | S | S | S |

Legenda:

S = Satisfatório

Para efeito de avaliação do erro amostral, o menor indicador entre (MAPE, sMAPE, $MEr$) foi o $MEr$.

O desvio padrão $s$ do $MEr$ dessa amostra foi de 5,80%.

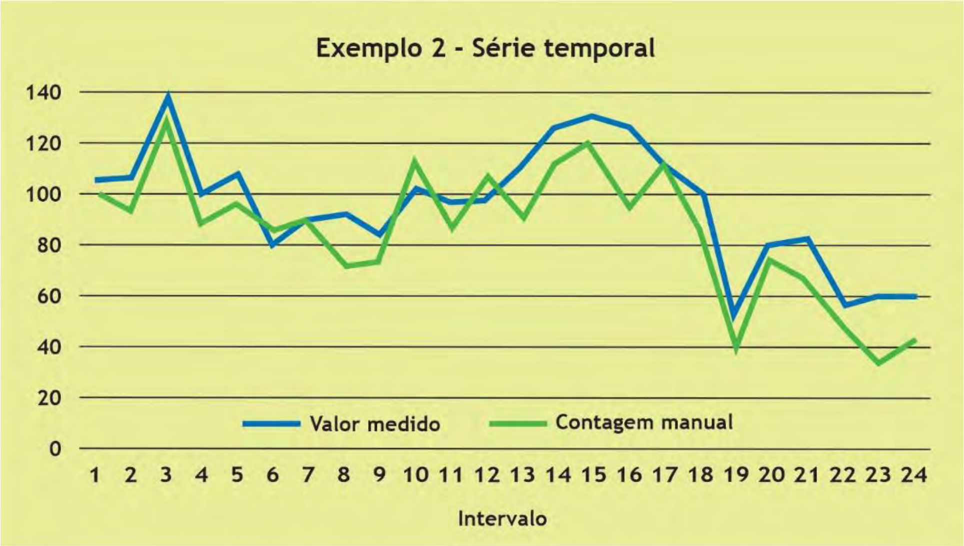
O $MEr$ e o seu erro amostral para essa amostra de 24 intervalos de 15 minutos foi de 6,50% e 2,45%, respectivamente, resultando em um intervalo de confiança para o $MEr$ de [4,05%; 8,95%]. Isto é:

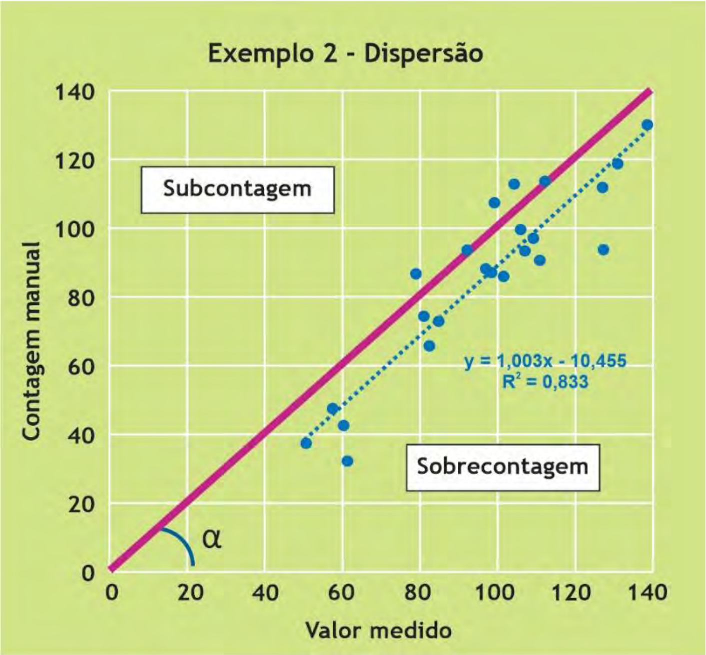
- $MEr$ estimado = 6,50%
- 4,05% ≤ $MEr$ verdadeiro ≤ 8,95%, com nível de confiança de 95%

Conclusão:

Apenas para efeitos dos valores adotados a título de ilustração, o sistema testado apresenta boa acurácia e boa precisão e pode ser considerado como satisfatório. O erro amostral de 2,45% pode ser reduzido aumentando o tamanho da amostra. Para 48 intervalos, o erro amostral é de 1,73%.

## 7.2 Exemplo 2




A análise gráfica mostra que os pontos azuis se afastaram um pouco da reta lilás, encontrando-se algo disperso em torno da linha azul pontilhada. Além disso, parece haver uma leve tendência de sobrecontagem. A linha azul pontilhada está quase paralela à reta lilás, o que sugere que talvez seja possível que os valores possam ser corrigidos por um fator de ajuste. Entretanto, para a aplicação de um fator de ajuste, além do fato de ser paralela à reta lilás, os pontos azuis devem estar próximos da reta azul pontilhada.

A análise visual dos gráficos não permite uma conclusão definitiva.

Quanto à análise dos indicadores:

| MAPE | WAPD | sMAPE | $MEr$ (médio) | $Er$ (total) | Min | | Correlação $r$ | | Análise gráfica | Avaliação final |
|---|---|---|---|---|---|---|---|---|---|---|
| **18,27%** | 14,64% | 15,86% | 14,13% | 10,58% | 10,58% | S | 91,27% | S | IN | S |

Legenda:

S = Satisfatório

IN = Inconclusivo

Para efeito de avaliação do erro amostral, o menor indicador entre (MAPE, sMAPE, $MEr$) foi o $MEr$.

O desvio padrão $s$ do $MEr$ foi de 9,99%.

O $MEr$ e o seu erro amostral para essa amostra de 24 intervalos de 15 minutos foi de 14,13% e 4,22%, respectivamente, resultando em um intervalo de confiança para o $MEr$ de [9,91%; 18,35%]. Isto é:

- $MEr$ estimado = 14,13%
- 9,91% ≤ $MEr$ verdadeiro ≤ 18,35%, com um nível de confiança de 95%

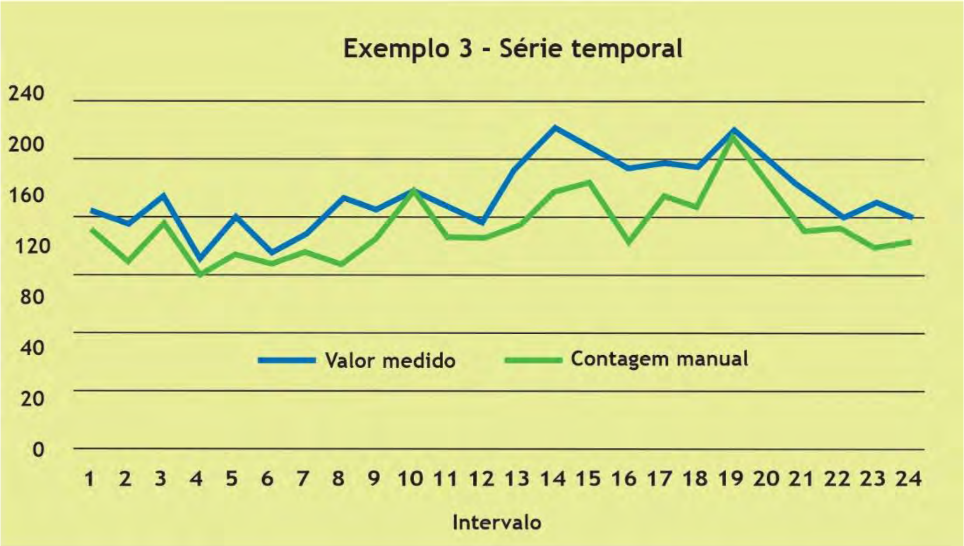
Conclusão:

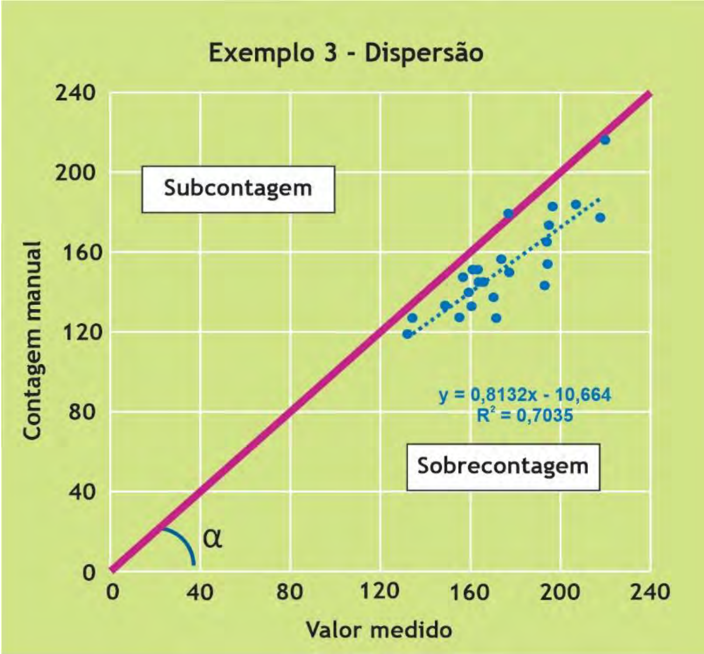
Embora a análise visual dos gráficos não tenha permitido chegar a um resultado conclusivo, todos os indicadores ficaram dentro dos parâmetros adotados a título de ilustração. Dessa forma, apenas para efeitos de exemplo, o sistema pode ser considerado como satisfatório.

Obviamente, se os valores estipulados para avaliação fossem mais exigentes do que aqueles adotados como ilustração: por exemplo, se $A_1 = 10\%$ (em vez de 17%), a conclusão seria pela repetição do teste ou mesmo pela rejeição, dependendo do valor de $A_2$.

Quanto ao erro amostral de 4,22%, ele ficou acima dos 2% admitido como máximo admissível. Além disso, o limite superior do intervalo de confiança de 18,35% é maior que $A_1 = 17\%$. Isso quer dizer que existe uma chance de o valor verdadeiro do $MEr$ ser maior que $A_1 = 17\%$. Com uma amostra de 48 intervalos, o erro amostral passa a ser de 2,98%, diminuindo a incerteza quanto ao valor verdadeiro do $MEr$. O ideal é que o tamanho da amostra seja tal que o erro amostral seja inferior a 2%.

## 7.3 Exemplo 3






Pela análise gráfica, percebe-se nitidamente que há uma sobrecontagem sistemática e que a reta de ajuste da regressão (linha azul pontilhada) não ficou paralela à reta lilás. Visualmente, os pontos azuis não ficaram tão afastados da reta azul pontilhada, apesar de o $R^2$ ter um valor relativamente baixo (0,7035).

A análise gráfica é inconclusiva.

Quanto à análise dos indicadores:

| MAPE | WAPD | sMAPE | $MEr$ (médio) | $Er$ (total) | Min | | Correlação $r$ | | Análise gráfica | Avaliação final |
|---|---|---|---|---|---|---|---|---|---|---|
| **14,98%** | 14,48% | 13,62% | 12,52% | 12,57% | 12,52% | S | 83,87% | RT | IN | RT |

Legenda:

S = Satisfatório
IN = Inconclusivo
RT = Repetição de teste

Para efeito de avaliação do erro amostral, o menor indicador entre (MAPE, sMAPE, $MEr$) foi o $MEr$.

O desvio padrão $s$ do $MEr$ foi de 6,70%.

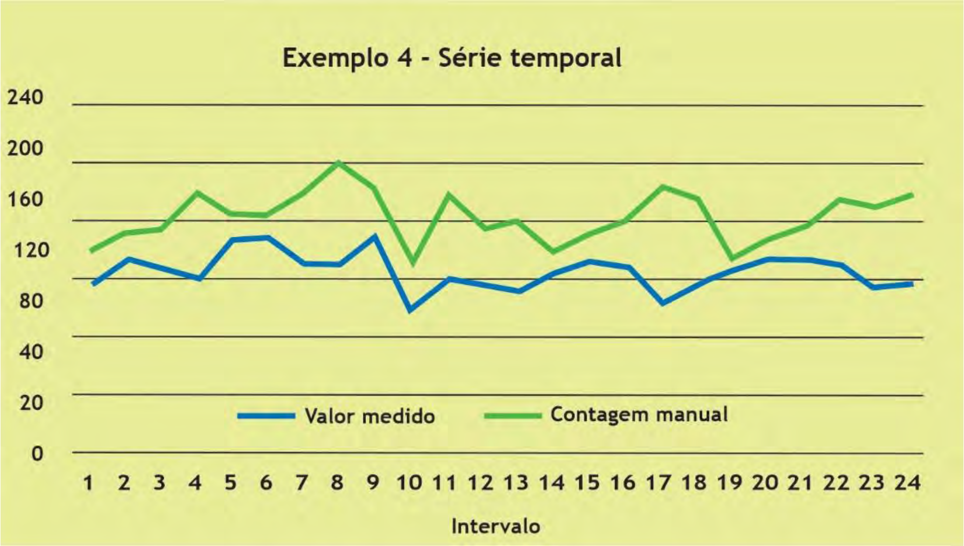
O $MEr$ e o seu erro amostral para essa amostra de 24 intervalos de 15 minutos foi de 12,52% e 2,83%, respectivamente, resultando em um intervalo de confiança para o $MEr$ de [9,69%; 15,35%]. Isto é:

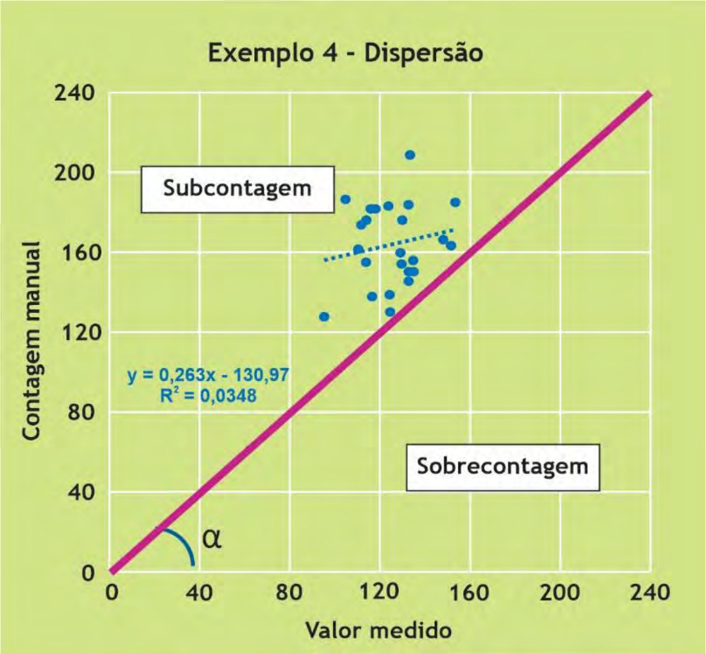
- $MEr$ estimado = 12,52%
- 9,69% ≤ $MEr$ verdadeiro ≤ 15,35%, com nível de confiança de 95%

Conclusão:

Embora a análise visual dos gráficos não permitiu um resultado conclusivo, o indicador de erro ficou dentro do parâmetro adotado a título de ilustração. Porém, a correlação ficou na faixa entre 80% a 90%. Dessa forma, apenas para efeitos de exemplo, a recomendação é de repetir o teste visando a correção dos erros sistemáticos que causaram a sobrecontagem.

## 7.4 Exemplo 4




Pela análise visual dos gráficos, percebe-se claramente que há uma subcontagem sistemática. Além disso, a reta azul pontilhada parece não ter relação com a reta lilás e os pontos se encontram dispersos de forma aleatória.

A análise gráfica mostra que os resultados não são satisfatórios.

Quanto à análise dos indicadores:

| MAPE | WAPD | sMAPE | *MEr* (médio) | *Er* (total) | Min | Correlação *r* | Análise gráfica | Avaliação final |
|---|---|---|---|---|---|---|---|---|
| **22,51%** | 23,26% | 26,11% | 22,51% | 23,26% | 22,51% NS | 18,64% NS | NS | NS |

Legenda:

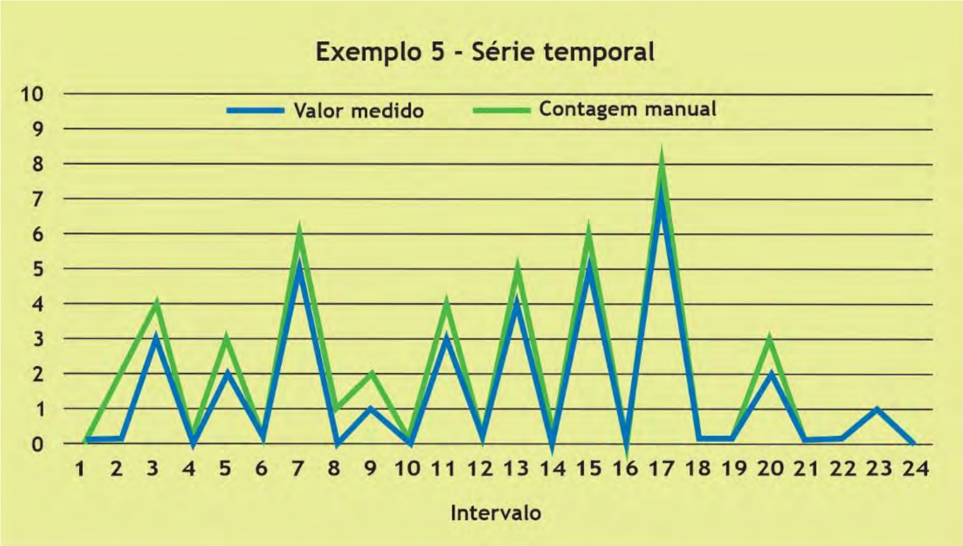
NS = Não satisfatório

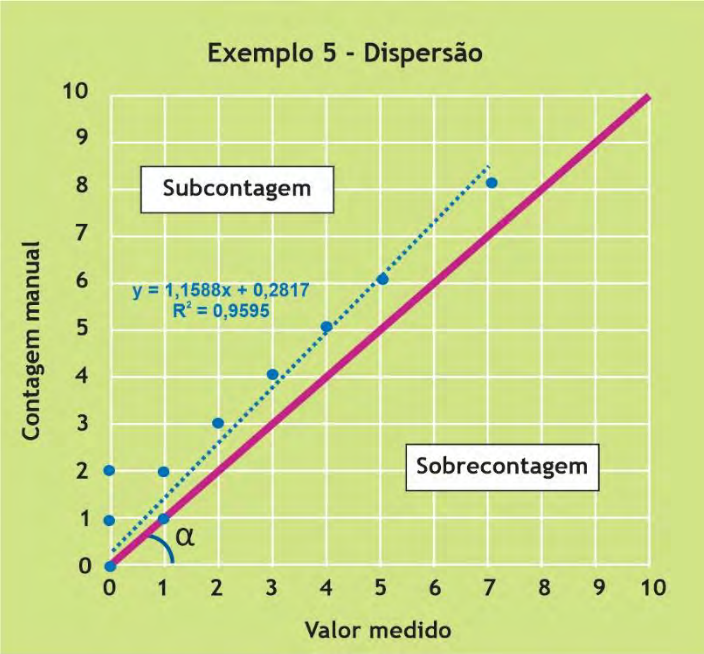
Todos os indicadores ficaram fora da faixa de valores estipulados como exemplo, corroborando com a análise gráfica.

Conclusão:

Tomando-se como referência os valores dos parâmetros estipulados como exemplo, o sistema é considerado não satisfatório.

## 7.5 Exemplo 5




Trata-se de uma contagem com baixíssimo fluxo (cerca de 6 veículos por hora). Pela análise visual dos gráficos, percebe-se claramente que há uma subcontagem sistemática, porém a reta azul pontilhada ficou próxima à reta lilás e os pontos azuis estão claramente relacionados com a reta azul pontilhada.

A análise gráfica não permite chegar a um resultado conclusivo.

Quanto à análise dos indicadores:

| MAPE | WAPD | sMAPE | $MEr$ (médio) | $Er$ (total) | Min | Correlação $r$ | Análise gráfica | Avaliação final |
|---|---|---|---|---|---|---|---|---|
| --- | 26,67% | 28,16% | 18,02% | 26,67% | 18,02% RT | 97,95% S | IN | RT |

Legenda:

S = Satisfatório
IN = Inconclusivo
RT = Repetição de teste

Por se tratar de uma contagem com fluxo muito baixo, houve vários intervalos em que o fluxo era nulo (valor verdadeiro = zero). Por isso, não foi possível calcular o MAPE.

Para efeito de avaliação do erro amostral, o menor indicador entre (MAPE, sMAPE, $MEr$) foi o $MEr$.

O desvio padrão $s$ do $MEr$ foi de 28,96%.

O $MEr$ e o seu erro amostral para essa amostra de 24 intervalos de 15 minutos foi de 18,02% e 12,23%, respectivamente, resultando em um intervalo de confiança para o $MEr$ de [5,79%; 30,25%]. Isto é:

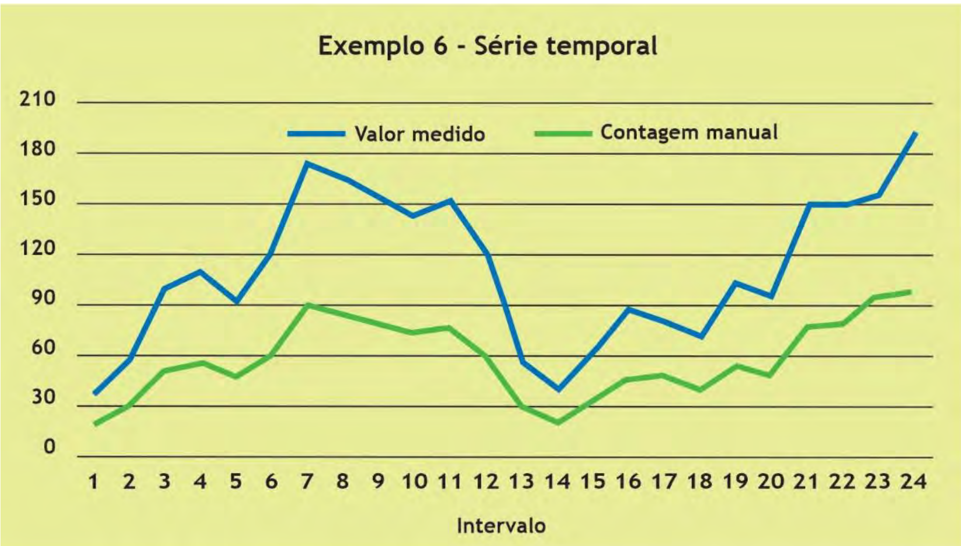
- $MEr$ estimado = 18,02%
- 5,79% ≤ $MEr$ verdadeiro ≤ 30,25%, com 95% de probabilidade

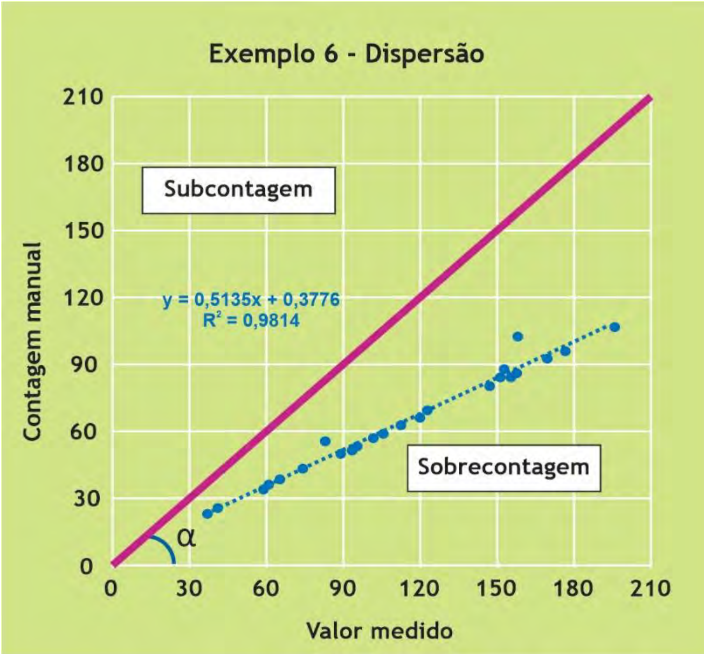
O erro amostral mostrou-se um valor elevado, com um intervalo de confiança muito amplo, gerando considerável incerteza com relação ao valor verdadeiro de $MEr$.

Conclusão:

Embora a análise visual dos gráficos não tenha permitido chegar a um resultado conclusivo, o indicador de erro ficou dentro da faixa de 17% a 20%, adotada a título de ilustração. A correlação ficou dentro do valor estipulado como exemplo. Dessa forma, apenas para efeitos de exemplo, a recomendação é de repetir o teste, com uma amostra maior, visando a correção dos erros sistemáticos que causaram a subcontagem, bem como diminuir o desvio padrão visando uma redução do erro amostral.

## Exemplo 6






Pela análise visual dos gráficos, percebe-se claramente que há uma sobrecontagem sistemática. Além disso, a reta azul pontilhada parece não ter relação com a reta lilás, porém os pontos azuis se encontram praticamente em cima da linha azul pontilhada, tanto que o $R^2$ está bem próximo de 1. Entretanto, os pontos azuis estão distantes da linha lilás, significando que os valores medidos estão afastados dos valores verdadeiros. Isso mostra que um $R^2$ próximo de 1 não significa necessariamente boa acurácia, significando apenas que o sistema tem boa precisão. Esse exemplo ilustra bem o caso da Figura 1-c: boa precisão, mas pouca acurácia.
A análise gráfica mostra que os resultados não são satisfatórios.

Quanto à análise dos indicadores:

| MAPE | WAPD | sMAPE | $MEr$ (médio) | $Er$ (total) | Min | Correlação $r$ | Análise gráfica | Avaliação final |
|---|---|---|---|---|---|---|---|---|
| **96,87%** | 96,07% | 62,44% | 49,10% | 49,00% | 49,00% NS | 99,07% S | NS | NS |

Legenda:

S = Satisfatório
NS = Não satisfatório

Apesar do alto índice de correlação, os indicadores de erro ficaram todos fora da faixa de valores estipulados como exemplo, sendo o sistema considerado não satisfatório, corroborando com a análise gráfica.

# BIBLIOGRAFIA CONSULTADA


[1] Feldman, Olga.
*The GEH Measure and Quality of the Highway Assignment Models*, 2012.

[2] Guido, Giuseppe; Gallelli, Vincenzo; Rogano, Daniele; Vitale, Alessandro.
*Evaluating the Accuracy of Vehicle Tracking Data Obtained from Unmanned Aerial Vehicles, International Journal of Transportation Science and Technology* 5 (2016) 136-151.

[3] Chowdhury, Mashrur; Khan, Sakib Mahmud; Khan, Zadid; Brunk, Katherine; Islam, Sababa; Keehan, McKenzie.
*Cost Effective Strategies for Estimating Statewide* AADT, *Final Report, February* 2019. FHWA-SC-18-10.

[4] Lessard, Adrien; Belisle, Francois; Bilodeau, Guillaume-Alexandre; Saunier, Nicolas.
*The Counting App, or How to Count Vehicles in* 500 *Hours of Video*.

[5] Chang, David K.; Saito, Mitsuru; Schultz, Grant G.; Eggett, Dennis L.
*Use of Hi-Resolution Data for Evaluating Accuracy of Traffic Volume Counts Collected by Microwave Sensors*.

[6] Puan, O. C.; Nor, N. S. M.; Mashros, N.; Hainin, M. R.
*Applicability of an Automatic Pneumatic-Tube-Based Traffic Counting Device for Collecting Data Under Mixed Traffic*, IOP *Conference Series: Earth and Environmental Science*, 2019.

[7] Weible, Benjamin.
*Investigating the Impact of Skewed Pneumatic Traffic-Counting Tubes on Accuracy, Annual Meeting* 2016.

[8] Nordback, Krista; Kothuri, Sirisha Murthy; Phillips, Taylor; Gorecki, Carson; Figliozzi, Miguel.
*Accuracy of Bicycle Counting with Pneumatic Tubes in Oregon*, 2016.

[9] Mendigorin, Lita; Peachman, John; White, Rod.
*The Collection of Classified Vehicle Counts in an Urban Area - Accuracy Issues and Results.*

[10] *Traffic Modelling Guidelines* – TfL: *Traffic Manager and Network Performance Best Practice - Version* 3.0

[11] Ryus, Paul; Ferguson, Erin; Laustsen, Kelly M.; Schneider, Robert J; Proulx, *Frank R; Hull, Tony; Luis Miranda-Moreno Methods and Technologies for Pedestrian and Bicycle Volume Data. Collection National Academies of Sciences, Engineering, and Medicine* (2014) – NCHRP *Project* 07-19.

[12] Ryus, Paul; Butsick, Andrew; Proulx, Frank R.; Schneider, Robert J.; Hull, Tony
*Methods and Technologies for Pedestrian and Bicycle Volume Data Collection: Phase* 2. *National Academies of Sciences, Engineering, and Medicine (2017)* – NCHRP 07-19(02) *Final Report*.

[13] *Highway Capacity Manual* – HCM 2010 – TRB

[14] Bratu, Mihaela
*The Reduction of Uncertainty in Making Decisions by Evaluating the Macroeconomic Forecasts Performance in Romania. Economic Research -* ***Ekonomska Istraživanja,*** Vol. 25 (2012) No. 2 (239-262).

[15] Kalaa, Mohamad Omar Al; Rajab, Samer; Refai, Hazem H.; Johnson, Daryl
*The Study of Vehicle Classification Equipment with Solutions to Improve Accuracy in Oklahoma, Final Report* - FHWA-OK-14-17 (2014).

[16] Richardson, Anthony J.; Ampt, Elizabeth S.; Meyburg, Arnim H.
*Survey Methods for Transport Planning*
Traffic_survey_form.pdf (no2hcf.co.uk)

[17] *Manual of Transportation Engineering Studies - Second Edition* – ITE

[18] Proulx, Frank R.; Schneider, Robert J.; Miranda-Moreno, Luis F.
*Performance Evaluation and Correction Functions for Automated Pedestrian and Bicycle Counting Technologies*
J. Transp. Eng. 2016.142

[19] Pesquisa e Levantamentos de Tráfego – Boletim Técnico 31 – CET (Partes 1 e 2)
Parte 1: http://www.cetsp.com.br/media/65280/bt31-%20pesquisa%20e%20levantamento%20de%20trafego-parte01.pdf
Parte 2: http://www.cetsp.com.br/media/69247/bt31-%20pesquisa%20e%20levantamento%20de%20trafego-parte02.pdf

[20] *What the MAPE Is Falsely Blamed for, Its True Weaknesses and Better Alternatives*!
https://www.statworx.com/de/blog/what-the-mape-is-falsely-blamed-for-its-true-weaknesses-and-better-alternat

[21] O que É Acurácia? Entenda o Conceito e Sua Importância.
https://blog.idwall.co/o-que-e-acuracia/

[22] Medição de Dados Experimentais, Incerteza e Propagação de Erro
www.fem.unicamp.br/~instmed/Incerteza_Old.htm

[23] Qual a Diferença Entre Precisão e Acurácia?
https://www.santiagoecintra.com.br/blog/geo-tecnologias/qual-a-diferenca-entre-precisao-e-acuraciay

[24] Precisão e Acurácia: Você Sabe a Diferença?
https://blog.cpetecnologia.com.br/precisao-e-acuracia-voce-sabe-a-diferenca/#:~:text=Entendendo

[25] Análise de Regressão: Como Interpretar o R-Quadrado e Avaliar a Qualidade de Ajuste?
Análise de regressão: Como interpretar o R-quadrado e avaliar a qualidade de ajuste? (minitab.com)

[26] Erros Aleatórios e Sistemáticos: O Que É Isso?
https://accmetrologia.com.br/erros-aleatorios-e-sistematico-o-que-e-isso/

[27] *Choosing the Correct Error Metric*: MAPE Vs. sMAPE – *The Pros and Cons of Two Popular Error Metrics*
https://towardsdatascience.com/choosing-the-correct-error-metric-mape-vs-smape-5328dec53fac

[28] Coeficientes de Correlação: Para que servem e como interpretá-los
Coeficientes de correlação: Para que servem e como interpreta-los? (operdata.com.br)