Paola Bongiovani

Mestrado: Engenharia de Sistemas Agrícolas (Agrometeorologia) - Universidade de São Paulo

Orientador: Dr. Paulo Cesar Sentelhas

Projeto: Mudanças climáticas e seus impactos na cultura da mandioca no Semiárido brasileiro e estratégias de manejo para mitigação das perdas

Áreas de interesse: agrometeorologia, modelagem (modelos de crescimento de culturas), mudanças climáticas.

Contato: paola.f.bongiovani@gmail.com / paola.bongiovani@usp.br

---

Link para exercícios

Exerc. Paola Bongiovani

Trabalho final

PROPOSTA A:

Análise estatística de dados observados x simulados com modelos

1. Contextualização

Um modelo de simulação é composto por dois ou mais modelos matemáticos, os quais consistem em um conjunto de equações que visam fornecer uma descrição matemática simplificada de um determinado fenômeno complexo do mundo real. A utilização de modelos de simulação tem sido cada vez mais difundida, pois eles nos permitem compreender de forma mais simples algo complexo da natureza, além de testar diferentes cenários e condições (de local, de solo, de clima, entre outros) e analisar padrões.

A partir de dados empíricos, inúmeras simulações são realizadas alterando-se os parâmetros ou coeficientes inerentes ao modelo, até que este esteja devidamente calibrado para os dados observados obtidos, ou seja, até que se atinja a melhor relação possível entre os dados simulados pelo modelo e os empíricos (ou observados). A análise dessa relação é realizada utilizando-se indicadores estatísticos, os quais devem ser recalculados a cada nova simulação, ou seja, a cada novo conjunto de dados simulados obtidos.

A performance do modelo em relação aos dados observados é verificado com base em indicadores estatísticos, como: coeficiente de determinação (r^2), índice de Pearson (r) (Tabela 1), índice de concordância (d) (WILLMOTT, 1982), índice de confiança ou de desempenho (c) (CAMARGO; SENTELHAS, 1997) (Tabela 2), o erro médio (EM), erro absoluto médio (EAM), raíz do erro quadrático médio (REQM) e eficiência do método utilizado ou simulação realizada (EF).

$$r^2 = \frac{\sum(S_i - \bar{O})^2}{\sum(O_i - \bar{O})^2}$$

$$r = \frac{\sum(O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum(O_i - \bar{O})^2 \sum(S_i - \bar{S})^2}}$$

$$d = 1 - \left[\frac{\sum(S_i - O_i)^2}{\sum(|S_i - \bar{O}| + |O_i - \bar{O}|)^2}\right]$$

$$c = d \cdot \sqrt{r^2}$$

$$EM = \frac{1}{n}\sum(S_i - O_i)$$

$$EAM = \frac{1}{n}\sum|S_i - O_i|$$

$$REQM = \frac{1}{n}\sqrt{\sum(S_i - O_i)^2}$$

$$EF = 1 - \left[\frac{\sum(O_i - \bar{O})^2 - \sum(O_i - S_i)^2}{\sum(O_i - \bar{O})^2}\right]$$

Em que Si corresponde aos valores simulados pelo modelo; Oi, aos valores observados; Ō é a média dos valores observados, n é o número total de valores observados ou estimados.

Tabela 1. Classificação dos valores do coeficiente de correlação de Pearson (r) (HOPKINS, 2000).

Coeficiente de correlação (r)	Classificação
0.0 a 0.1	Muito baixa
0.1 a 0.3	Baixa
0.3 a 0.5	Moderada
0.5 a 0.7	Alta
0.7 a 0.9	Muito alta
0.9 a 1.0	Quase perfeita

Tabela 2. Critério de interpretação do desempenho do modelo, pelo índice "c" de Camargo e Sentelhas (1997).

Índice de desempenho "c"	Desempenho
>0.85	Ótimo
0.76 a 0.85	Muito bom
0.66 a 0.75	Bom
0.61 a 0.65	Mediano
0.51 a 0.60	Sofrível
0.41 a 0.50	Mau
≤0.40	Péssimo

Nesse contexto, destaca-se a necessidade de uma função que, a partir de um conjunto de dados observados e vários conjuntos de dados provenientes de simulações com modelos, realize todos os cálculos de indicadores estatísticos automaticamente - para analisarmos a performance dos modelos de modo mais prático e eficiente -, e que também crie um gráfico com os dados observados e a cada conjunto de dados simulados - de modo a permitir uma melhor visualização da relação entre eles. A função vai pegar um conjunto de dados observados e diferentes conjuntos de dados simulados (sendo um conjunto de dados simulados em cada coluna de uma planilha) e realizar todos os cálculos estatísticos e análises dos indicadores que apresentam classificações, simultaneamente, gerando uma tabela com os resultados dos indicadores estatísticos, além de um gráfico de dispersão dos dados observados por cada grupo de dados simulados, com a reta e equação da regressão linear, além de uma reta 1:1.

2. Planejamento da função

Entrada:

obs.sim(`obs`, `sim`, `indice`)

- `obs` = conjunto de dados observados (classe:numeric, obs ≥ 0);
- `sim` = conjuntos de dados simulados, cada conjunto em uma coluna diferente (classe: data.frame, sim ≥ 0);
- `indice` = índice(s) estatístico(s) desejado(s) (classe: fator);

Verificando os parâmetros:

- `obs` é um número e é maior ou igual a zero? Se não, escreve: "obs precisa ser número e ≥ 0."
- `sim` é um número e é maior ou igual a zero? Se não, escreve: "sim precisa ser da classe data.frame e ≥ 0."
- `obs` e `sim` apresentam mesma quantidade de dados? Se não, escreve: "obs e sim devem apresentar mesma quantidade de dados."
- `indice` é diferente de `all`,`r2`, `r`, `d`, `c`, `EM`, `EAM`, `REQM` ou `EF`? Se for, escreve: "indice pode ser

all, r2, r, d, c, EM, EAM, REQM ou EF."

Pseudocódigo:

1. Cria objeto `obs` com `obs`;
2. Cria data.frame `sim` com `sim`;
3. Cria um objeto com resultado dos índice estatísticos selecionados:
    1. Calcula o r^2 e guarda em `r2`;
    2. Calcula o r e guarda em `r`,
    3. Calcula o d e guarda no `d`;
    4. Calcula o c e guarda no `c`;
    5. Calcula o erro médio e guarda no `EM`;
    6. Calcula o erro absoluto médio e guarda no `EAM`;
    7. Calcula a raíz do erro quadrático médio e guarda no `REQM`;
    8. Calcula a eficiência da simulação e guarda em `EF`;
4. Cria data frame com `r2`, `r` e a classificação da correlação, `d`, `c` e a interpretação do desempenho, `EM`, `EAM`, `REQM`, `EF`, sendo uma linha para cada coluna de `sim`;
5. Cria objeto `xy` com a regressão linear de `obs` com cada coluna de `sim`;
6. Cria gráfico de dispersão com `obs` (eixo x) e cada coluna de `sim` (eixo y), com a reta da regressão linear, a equação da reta, o r^2 e uma reta 1:1.

Saída:

- Gráfico de dispersão dos dados observados e simulados pelo modelo, com a reta obtida com a regressão linear, o r^2 e uma reta 1:1;
- Data frame com os resultados dos índices estatísticos escolhidos, sendo uma linha de resultados para cada coluna de dados simulados.

3. Referências

CAMARGO, A. P.; SENTELHAS, P. C. Avaliação do desempenho de diferentes métodos de estimativa da evapotranspiração potencial no estado de São Paulo. Revista Brasileira de Agrometeorologia, v. 5, n. 1, p. 89-97, 1997.

HOPKINS, W. G. Correlation coefficient: a new view of statistics, 2000.

WILLMOTT, C. J. Some comments on the evaluation of model performance. Bulletin of the American Meteorological Society, v. 63, n. 11, p. 1309–1313, 1982.

PROPOSTA B:

Estimativa do álcool no sangue e efeitos no corpo e condução veicular

1. Contextualização

O consumo de álcool é algo presente em nosso dia-a-dia. Com a criação da Lei Seca no Brasil, houve um aumento de conscientização sobre os perigos do consumo de álcool, em geral, e seus efeitos em

nossas ações e reações à determinados estímulos, como os relacionados à condução veicular. A proposta de “Tolerância Zero” trazida pela Lei foi uma estratégia que pune, basicamente, qualquer usuário que tenha ingerido bebidas alcóolicas. Por consequência, há pessoas que deixam de ingerir álcool por medo de pagar multas ou perder a CNH (e se tornam “o motorista da vez”), e outras que consomem e montam estratégias para evitar caminhos com possibilidade de haver barreiras policiais. Entretanto, nem sempre estamos cientes dos efeitos causados em nosso organismo de acordo com a quantidade de álcool que ingerimos, sejam elas maiores ou menores.

Widmark (1981) criou uma fórmula para determinar a quantidade de álcool no sangue (c) a partir de cálculos que demandam três informações: quantidade de álcool ingerida (A), peso corporal do consumidor (p) e o coeficiente de redução (r, cujos valores médios variam de acordo com o gênero). A quantidade de álcool ingerida, em grama de álcool por quilo de álcool, é obtida a partir dos teores alcóolicos das bebidas consumidas (em %), o volume consumido (em ml) e o peso específico (ou densidade) do álcool, padronizado em 0.79 gramas.

$$A = p \times r \times c$$

$$A = teor\ alcoólico \times volume \times densidade\ do\ álcool$$

A função proposta receberia, então, o gênero e peso do consumidor, o tipo de bebida ingerida e o volume ingerido, e faria cálculos de acordo com os valores de entrada utilizando-se dados como os exemplificados na tabela abaixo:

Bebida	Dose padrão (ml)	Teor alcoólico (%)
Chopp pequeno	250	5
Chopp	350	5
Cerveja - Lata grande	450	5
Cerveja - Lata pequena	350	5
Cerveja - 600	600	5
Cerveja - Litro	1000	5
Ice	330	5.5
Vinho branco	140	10
Champagne	140	11
Vinho tinto	140	12.5
Vinho tinto - garrafa	750	12.5
Saquê	100	15.5
Licor	80	28
Cachaça de jambú	40	38
Tequila	40	38
Vodka	40	40
Uísque	225	40
Caipirinha	37.5	40
Caipirinha com chorinho	50	40
Cachaça	40	40

Para concordar com o sistema atualizado no Brasil, em que a Taxa de Álcool no Sangue (TAS) é obtida em gramas por litro de sangue (g/L ou dg/L), a função converteria automaticamente o resultado obtido na função de Widmark (1981), utilizando-se do fator de conversão de Simonin (1982). Em seguida, a quantidade de álcool no sangue seria utilizada para verificar quais os seus eventuais efeitos no corpo do consumidor, além dos possíveis efeitos e riscos de acidente ao conduzir um

veículo.

2. Planejamento da função

Entrada:

drink.n.drive(gen, p, beb, vol)

- gen = gênero do consumidor (classe: fator, gen = M ou F);
- p = peso do consumidor, em kg (classe: numeric, p > 0);
- beb = tipo(s) de bebida(s) consumida(s) (classe: fator);
- vol = volume(s) consumido(s) de (cada) bebida, em ml (classe: integer, vol > 0).

Verificando os parâmetros:

- gen é um fator? Se não, escreve: “gen precisa ser um fator.”
- p está em kg e é um número maior que zero? Se não, escreve: “p precisa estar em kg, ser um número e ser > 0.”
- beb é um fator? Se não, escreve: “beb deve ser um fator.”
- vol está em ml e é um número inteiro e maior que zero? Se não, escreve: “vol deve estar em ml, ser um número inteiro e ser > 0.”

Pseudocódigo:

1. Cria data frame fem se gen=F ou masc se gen=M, com respectivo r;
2. Calcula A com vol e a lista beb e guarda no fem ou masc;
3. Calcula c com A, p e r;
4. Converte unidade de c;
5. Seleciona no body.ef os valores para c e guarda em fem ou masc;
6. Seleciona no cond.ef os valores para c e guarda em fem ou masc;
7. Recupera data frame com a estimativa de c e os efeitos no corpo e condução.

Saída:

- Data frame com a estimativa da quantidade de álcool no sangue do consumidor, os efeitos da quantidade no corpo e os efeitos e/ou riscos de acidente na condução.

3. Referências

SIMONIN, C. Medicina legal judicial. Barcelona: Editorial JIMS, 1962.

WIDMARK, E. M. P. Principle and applications of medicolegal alcohol determination. California (USA): Biomedical Publications, 1981.

> — *Alexandre Adalardo de Oliveira 2018/05/14 11:10* Oi Paola,
>
> **A proposta 1** é interessante, mas simples de implementar, já que o acoplamento entre dados e simulação é um cálculo simples dos índices apresentados. Deve incorporar algo mais para que a função seja um desafio na implementação da linguagem. Uma sugestão é que os dados simulados não seja apenas uma simulação, mas um conjunto de simulações

em um data.frame. Algo muito comum em modelos computacionais. Nesse sentido, vc poderia fazer uma distribuição dos índices e apresentá-los não apenas como um valor médio, mas a média (de todas as simulações) e o intervalo de confiança.

A proposta 2 é interessante tb. mas apresenta o mesmo problema da primeira, a sua implementação, como apresentada, é muito simples. Minha sugestão é fazer o calculo não para uma pessoa, mas para um grupo de pessoas e retornar o valor para esse grupo. Algumas coisas precisam ser melhor explicadas, por exemplo “Em seguida, a quantidade de álcool no sangue seria utilizada para verificar quais os seus eventuais efeitos no corpo do consumidor, além dos possíveis efeitos e riscos de acidente ao conduzir um veículo”... essa frase sugere que a função fará muito mais do que é apresentado no psedocodigo, mas essa parte não está especificada.

Minha sugestão é que atualize a proposta 1 com as sugestões, mas caso queira ir pela 2, não há problema, mas precisa trabalhar melhor a descrição. Aguardamos seu retorno para fechar a propostas.

Proposta final: A

Links para acesso:

- Link para a página função obs.sim: Função obs.sim
- Link para a página do help da função obs.sim: HELP