

Dois casos de controlo estatístico da concentração de poluentes

Descrevem-se dois casos fictícios que tratam o caso de controlo de poluentes e o tratamento estatístico das leituras envolvendo prémios, isto é, formas pecuniárias de premiar o sucesso na manutenção de valores baixos e de penalizar o insucesso nessa manutenção. O resultado destes prémios pode ser canalizado para obras sociais dentro das empresas em que ocorrem. Por esta via, a empresa pode constatar a redução significativa da frequência dos incidentes “fora de controlo” e os trabalhadores ganham um estímulo para um melhor desempenho.

Os casos aqui retratados, embora fictícios, existem em instalações técnicas variadas nas quais o controlo das variáveis de *output*, embora com o apoio de alguma automação, dependem sobretudo da competência e organização de equipas de operadores. É o caso, por exemplo, das ETAR (Estação de Tratamento de Águas Residuais) ou das ETA (Estação de Tratamento de Água), públicas ou privadas.

Caso 1

Um processo fabril contínuo rejeita cobre residual no efluente líquido, cuja concentração depende do cuidado posto no controlo daquele processo por parte do pessoal de cada turno. A monitorização deste poluente é realizada em momentos aleatórios dentro de cada hora ao longo de cada dia. O próximo Quadro mostra os resultados das leituras do dia de ontem.

Hora da leitura	p.p.m.	Hora da leitura	p.p.m.	Hora da leitura	p.p.m.
0.47	55	8.44	52	16.54	72
2.00	52	9.50	58	18.01	44
3.10	64	10.42	54	18.57	44
4.14	53	11.36	63	19.49	58
5.01	56	12.35	34	20.55	60
5.52	56	13.21	56	22.06	54
6.42	50	14.24	45	22.55	53
7.54	45	15.39	62	23.41	54

Quadro 1 – Concentrações observadas ao longo de 1 dia

- Existe algum valor atípico (singular ou *outlier*)? Em caso afirmativo, quais são e que cuidados devem merecer tal constatação?
- Qual é a probabilidade de esta concentração se situar entre 48 e 58 p.p.m.?
- A empresa impôs-se a seguinte regra: Concentrações iguais ou inferiores a 47 p.p.m. (LICA – Limite Inferior da Concentração Aceitável) são premiadas com um bónus para acções sociais dos trabalhadores no montante de 10 € por cada p.p.m. de diferença. Concentrações iguais ou superiores a 61 p.p.m. (LSCA – Limite Superior de Concentração Aceitável) são penalizadas com uma multa no montante de 5 € por cada p.p.m. de diferença. Qual o valor do prémio (bónus - multas) correspondente ao dia de ontem?
- Qual o valor da concentração C^* que apresenta uma probabilidade de ser menor ou igual a 0,70? $P(C \leq C^*) = 0,7 \rightarrow C^* = ?$
- Construa um intervalo de confiança para o valor médio com um nível de confiança de 0,98 e diga como o interpreta.
- Se pretendermos um erro não superior a 2 p.p.m. para um nível de confiança de 0,98, quantas mais observações seriam necessárias realizar?
- Qual o valor esperado do prémio num qualquer dia? Suponha que a amostra do dia de ontem é representativa dos resultados obtidos em cada dia.

Resolução (usando o EXCEL)

- a) $Q1 = \text{PERCENTILE}(\text{Dados}; 0,25) = 51,5$
 $Q3 = \text{PERCENTILE}(\text{Dados}; 0,75) = 58$
 $LIA = Q1 - 1,5 \times (Q3 - Q1) = 51,5 - 1,5 \times (58 - 51,5) = 41,75$
 $LSA = Q3 + 1,5 \times (Q3 - Q1) = 58 + 1,5 \times (58 - 51,5) = 67,75$

Procurando no Quadro anterior, encontramos os *outliers*: 34 e 72

Dever-se-á inquirir a razoabilidade destes valores, pois podem dever-se a erros de anotação ou a circunstâncias normais, embora pouco frequentes. No primeiro caso, deverão ser eliminados da análise. No segundo caso, deverão ser considerados.

- b) Ordenando as três colunas cinza do Quadro 1 por ordem crescente (ou decrescente), obtemos: $P(C \leq 58) - P(C \leq 48) = \text{FREQUENCY}(\text{Dados ordenados}; 58) - \text{FREQUENCY}(\text{Dados ordenados}; 48) = 0,792 - 0,208 = 0,584$ (ou 14 em 24)
- c) Uma vez os dados ordenados, por exemplo, por ordem crescente, obtemos o Quadro 2.

<i>n</i>	Dados ordenados g/m ³	Prémio €
1	34	130
2	44	30
3	44	30
4	45	20
5	45	20
6	50	0
7	52	0
8	52	0
9	53	0
10	53	0
11	54	0
12	54	0
13	54	0
14	55	0
15	56	0
16	56	0
17	56	0
18	58	0
19	58	0
20	60	0
21	62	-5
22	63	-10
23	64	-15
24	72	-55

Quadro 2 – Dados ordenados e prémios calculados

Neste Quadro, por exemplo, os valores de ordem 5 e de ordem 22 que figuram na terceira coluna, foram calculados da seguinte forma, respectivamente:

Ordem 5: Como $45 < 47 \text{ g/m}^3$: $(47 - 45) \times 10 = 20 \text{ €}$

Ordem 22: Como $63 > 61 \text{ g/m}^3$: $(63 - 61) \times (-5) = -10 \text{ €}$

A soma da terceira coluna fornece o resultado: Prémio devido = 145 €

- d) $P(C \leq C^*) = 0,7 \rightarrow C^* = \text{PERCENTILE}(\text{Dados}; 0,7) = 56,2 \text{ p.p.m.}$

- e) Sendo a média dos dados $\bar{x} = 53,92 \text{ g/m}^3$ e o desvio padrão $s = 7,91 \text{ g/m}^3$, o semi-intervalo de confiança ε será:

$$\varepsilon = t_{\alpha/2} \cdot s / \sqrt{n} = 2,5 \times 7,91 / \sqrt{24} = 4,036 \text{ p.p.m.}$$

Logo, o intervalo de confiança será $(53,92 - 4,036 \leq \mu \leq 53,92 + 4,036)$ ou $(49,88 \leq \mu \leq 57,95)$, ou seja, a probabilidade da média da população μ , da qual esta amostra foi retirada, se encontrar dentro deste intervalo é 0,98: $P(49,88 \leq \mu \leq 57,95) = 0,98$

- f) O número total de observações necessárias realizar é dado por: $n = (t_{\alpha/2})^2 \cdot (s)^2 / \varepsilon^2 = (2,50)^2 \cdot (7,91)^2 / (2)^2 = 98$. Mas como já foram realizadas 24 observações, restam em falta $(98 - 24) = 74$ observações.
- g) Os parâmetros da distribuição da população de onde esta amostra proveio foram estimados e resultaram:

Média (igual à da amostra) = AVERAGE(Dados) = 53,92 p.p.m.

Desvio padrão = STDEV.P(Dados) = 7,74 p.p.m.

O teste de KS (Kolmogorov Smirnov) para um nível de significância $\alpha = 0,05$ evidenciou um ajustamento dos dados da amostra a uma distribuição Normal com D observado (D) = 0,03831 < D crítico (D_{α}) = 0,26931, pelo que passamos a considerar esta distribuição teórica nos cálculos posteriores em lugar dos dados empíricos.

Simulamos agora 24 valores aleatórios da função de distribuição Normal com aqueles parâmetros. Cada um referente a um momento aleatório dentro de cada uma das 24 horas de 1 dia. Calculamos depois para cada leitura obtida o correspondente prémio – umas vezes um bónus, outras vezes uma multa. O Quadro 3 mostra um exemplo resultante de uma iteração.

Ordem	Leituras Simuladas (LS)	Iterações
1	40,17624712	68,24
2	54,03071951	0,00
3	68,67350985	-38,37
4	43,09542923	39,05
5	59,57857219	0,00
6	61,71405974	-3,57
7	59,53897337	0,00
8	56,51165539	0,00
9	41,13189342	58,68
10	53,18593812	0,00
11	51,99430613	0,00
12	48,09311741	0,00
13	45,71926656	12,81
14	53,63464779	0,00
15	63,65829067	-13,29
16	52,25055172	0,00
17	52,27101079	0,00
18	51,04364407	0,00
19	48,38975406	0,00
20	47,70021036	0,00
21	48,34783394	0,00
22	53,75058039	0,00
23	51,10824516	0,00
24	56,91248412	0,00

Quadro 3 – Exemplo de 24 leituras aleatórios num dia e cálculo do correspondente prémio

A segunda coluna resulta da Expressão: =NORMINV(RAND()); Média; Desvio Padrão)

A terceira coluna resulta da Expressão: =IF(LS <= LICA;(LICA - LS)*Bónus; IF(LS > LSCA;(LS - LSCA)*Multa;0))

A soma da terceira coluna resulta igual a 123,54 €; importância que constitui o prémio instantâneo obtido nesta iteração em particular.

Para criar representatividade estatística, temos de repetir este resultado centenas ou mesmo milhares de vezes. Usando a aplicação EXCEL “Repetidor” que acompanha os meus dois últimos livros, concluímos após, por exemplo, 1.000 corridas:

Prémio esperado = 98,6 €/dia com um erro amostral de +/- 8,36 €/dia. Uma maior precisão requererá um maior número de corridas (repetições).

Caso 2

A concentração desejável de um determinado componente químico num processo industrial situa-se num intervalo entre 80 e 110 g/m³. Este teor é controlado manualmente. As leituras realizadas durante vários dias revelaram que o teor seguia uma distribuição de frequência Normal centrada em 100 g/m³ com um desvio padrão de 20 g/m³.

Na empresa encontra-se em vigor um prémio consistindo num bónus de 250 €/dia quando aquele intervalo é respeitado e numa multa de 150 €/dia quando não o é. Para efeito de cálculo do prémio consideram-se só a parte inteira das medições. Assim, se a medição for, por exemplo, 79,3 g/m³ ou 79,7 g/m³, consideramos o resultado como sendo 79 g/m³.

Nestas circunstâncias, qual o valor esperado do prémio diário (bónus - multa)?

Resolução (usando o EXCEL)

$$P(\text{observações} \leq 79) = \text{NORMDIST}(79; 100; 20; 1) = 0,146859$$

$$P(\text{observações} \leq 110) = \text{NORMDIST}(110; 100; 20; 1) = 0,691462$$

$$\text{Prémio} = (0,691462 - 0,146859) \times 250 + [1 - (0,691462 - 0,146859)] \times (-150) \cong 68 \text{ €/dia}$$

Rui Assis
18 Setembro 2018
rassis46@gmail.com
www.rassis.com