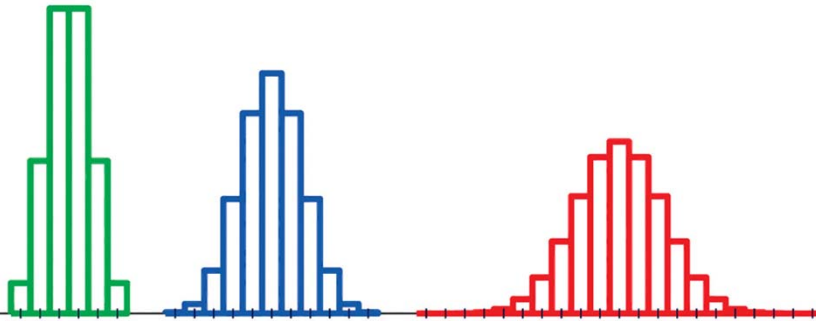


Representatividade da amostragem ambiental

25 de Outubro de 2019



Ricardo Bettencourt da Silva
[rjsilva@fc.ul.pt]

1

Índice

1. Análise de variâncias

1.1 ANOVA de um factor

1.2 A aritmética da ANOVA

2. Método Monte Carlo

2

1. Análise de variâncias

A análise de variâncias é uma ferramenta muito útil para comparar conjuntos de médias.

Exemplos: 1) Comparação de condições de conservação de amostras;
2) Distinguir a variabilidade de um sistema da precisão da repetibilidade da análise de amostras do sistema.

Nestes estudos, existem sempre **dois tipos de fontes de variação**, os **erros aleatórios** e o designado **factor controlado** (exemplo 1: condições de conservação).

A análise de variâncias (ANOVA) é uma ferramenta estatística usada para separar e estimar as diferentes causas de variação:

Varição causada pela alteração do factor controlado
Vs.
Varição causada pelos erros aleatórios

A ANOVA testa se a alteração do factor controlado produz diferenças significativas nas médias obtidas.

3

1. Análise de variâncias

A ANOVA pode ser usada para distinguir fatores aleatórios (ex.: repetibilidade das determinações + heterogeneidade da amostra).

Quando é estudado um factor controlado utiliza-se uma “ANOVA de um factor (one-way ANOVA)”.

Podem ser estudados múltiplos factores controlados recorrendo a outras ferramentas estatísticas.

4

1. Análise de variâncias

1.1 ANOVA de um factor

Habitualmente os dados estudados através de uma ANOVA de um factor são apresentados em forma de tabela:

							Média
Amostra 1	x_{11}	x_{12}	x_{1j}	x_{1n}	\bar{x}_1
Amostra 2	x_{21}	x_{22}	x_{2j}	x_{2n}	\bar{x}_2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Amostra i	x_{i1}	x_{i2}	x_{ij}	x_{in}	\bar{x}_i
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Amostra h	x_{h1}	x_{h2}	x_{hj}	x_{hn}	\bar{x}_h
	Média global						\bar{x}

A ANOVA testa se a diferença entre as médias das amostras é demasiado grande para ser explicada pelos erros aleatórios.

5

1. Análise de variâncias

1.1 ANOVA de um factor

A média, \bar{x} , das médias das diversas amostras ($\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$) é igual à média de todos os resultados x_{ij} .

							Média
Amostra 1	x_{11}	x_{12}	x_{1j}	x_{1n}	\bar{x}_1
Amostra 2	x_{21}	x_{22}	x_{2j}	x_{2n}	\bar{x}_2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Amostra i	x_{i1}	x_{i2}	x_{ij}	x_{in}	\bar{x}_i
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Amostra h	x_{h1}	x_{h2}	x_{hj}	x_{hn}	\bar{x}_h
	Média global						\bar{x}

H_0 : Todas as amostras têm como origem a mesma população com média μ e variância σ^2 .

Considerando H_0 , σ^2 pode ser estimado de duas maneiras: através da variação intra-amostra ou da variação inter-amostra (entre amostras).

6

1. Análise de variâncias

1.1 ANOVA de um factor

Varição intra-amostra:

							Média
Amostra 1	x_{11}	x_{12}	x_{1j}	x_{1n}	\bar{x}_1
Amostra 2	x_{21}	x_{22}	x_{2j}	x_{2n}	\bar{x}_2
⋮	⋮	⋮		⋮		⋮	⋮
Amostra i	x_{i1}	x_{i2}	x_{ij}	x_{in}	\bar{x}_i
⋮	⋮	⋮		⋮		⋮	⋮
Amostra h	x_{h1}	x_{h2}	x_{hj}	x_{hn}	\bar{x}_h
							Média global \bar{x}

$s_1^2 = \sum_j (x_{1j} - \bar{x}_1)^2 / n - 1$

A média das variâncias s_i^2 é uma estimativa de σ^2 : $s^2 = \sum_i s_i^2 / h$

Esta estimativa tem $[h \times (n-1)]$ graus de liberdade.

Importante: Esta estimativa não é função das médias das amostras!

A formula geral usada para estimar a variância intra-amostra é:

$$\sigma^2 = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 / h(n-1)$$

← Média dos quadrados

7

1. Análise de variâncias

1.1 ANOVA de um factor

Varição inter-amostra:

Se as amostras forem originárias da mesma população (μ, σ^2), as suas médias \bar{x}_i são originárias de uma população com variância σ^2/n . Assim, se H_0 for verdadeira, a variância da média das amostras é uma estimativa de σ^2/n .

							Média
Amostra 1	x_{11}	x_{12}	x_{1j}	x_{1n}	\bar{x}_1
Amostra 2	x_{21}	x_{22}	x_{2j}	x_{2n}	\bar{x}_2
⋮	⋮	⋮		⋮		⋮	⋮
Amostra i	x_{i1}	x_{i2}	x_{ij}	x_{in}	\bar{x}_i
⋮	⋮	⋮		⋮		⋮	⋮
Amostra h	x_{h1}	x_{h2}	x_{hj}	x_{hn}	\bar{x}_h
							Média global \bar{x}

Variância das médias

$$\sum_i (\bar{x}_i - \bar{x})^2 / (h-1)$$

Logo o estimativa “inter-amostra” de σ^2 é calculada pela equação (desvio padrão das médias $\times n$):

$$\sigma^2 = n \sum_i (\bar{x}_i - \bar{x})^2 / (h-1)$$

Esta estimativa está associada a $(h-1)$ graus de liberdade.

8

1. Análise de variâncias

1.1 ANOVA de um factor

Varição inter-amostra:

Importante: A estimativa de σ obtida com base na variação “inter-amostras”:

$$\sigma^2 = n \sum_i (\bar{x}_i - \bar{x})^2 / (h-1)$$

... não é função da variabilidade “intra-amostras” visto que se baseia nas médias das amostras.

9

1. Análise de variâncias

1.1 ANOVA de um factor

Comparação das duas estimativas de σ :

Se H_0 for verdadeira, as estimativas de σ baseadas na variabilidade “intra” e “inter-amostras” não são significativamente diferentes. Se H_0 não for correcta a estimativa de σ baseada na variação “inter-amostras” é superior à estimativa de σ baseada na variação “intra-amostras”.

Desta forma, as duas estimativas de σ são comparados através de um teste-F unilateral:

$$F_{cal.} = \frac{n \sum_i (\bar{x}_i - \bar{x})^2 / (h-1)}{\sum_i \sum_j (x_{ij} - \bar{x}_i)^2 / [h(n-1)]}$$

Este F calculado é comparado com o F tabelado:

$$F_{tab.}(h-1; h(n-1); P; unilateral)$$

Se $F_{cal.} > F_{tab.}$, H_0 é rejeitada.

10

1. Análise de variâncias

1.1 ANOVA de um factor

Valores críticos de F (P = 0,05 para testes unilaterais; P = 0,1 para testes bilaterais)													
v_2/v_1	1	2	3	4	5	6	7	8	9	10	12	15	20
1	161,44	199,49	215,70	224,58	230,16	233,98	236,76	238,88	240,54	241,88	243,90	245,94	248,01
2	18,513	19,000	19,164	19,247	19,296	19,329	19,353	19,371	19,385	19,396	19,412	19,429	19,446
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,785	8,745	8,703	8,660
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,912	5,858	5,803
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,678	4,619	4,558
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	4,000	3,938	3,874
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	3,575	3,511	3,445
8	5,318	4,459	4,066	3,838	3,688	3,581	3,500	3,438	3,388	3,347	3,284	3,218	3,150
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,073	3,006	2,936
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,913	2,845	2,774
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	2,788	2,719	2,646
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	2,687	2,617	2,544
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671	2,604	2,533	2,459
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602	2,534	2,463	2,388
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544	2,475	2,403	2,328
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538	2,494	2,425	2,352	2,276
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494	2,450	2,381	2,308	2,230
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456	2,412	2,342	2,269	2,191
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423	2,378	2,308	2,234	2,155
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348	2,278	2,203	2,124

11

1. Análise de variâncias

1.1 ANOVA de um factor

Valores críticos de F (P = 0,05 para testes unilaterais; P = 0,1 para testes bilaterais)													
v_2/v_1	1	2	3	4	5	6	7	8	9	10	12	15	20
1	161,44	199,49	215,70	224,58	230,16	233,98	236,76	238,88	240,54	241,88	243,90	245,94	248,01
2	18,513	19,000	19,164	19,247	19,296	19,329	19,353	19,371	19,385	19,396	19,412	19,429	19,446
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,785	8,745	8,703	8,660
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,912	5,858	5,803
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,678	4,619	4,558
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	4,000	3,938	3,874
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	3,575	3,511	3,445
8	5,318	4,459	4,066	3,838	3,688	3,581	3,500	3,438	3,388	3,347	3,284	3,218	3,150
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,073	3,006	2,936
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,913	2,845	2,774
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	2,788	2,719	2,646
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	2,687	2,617	2,544
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671	2,604	2,533	2,459
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602	2,534	2,463	2,388
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544	2,475	2,403	2,328
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538	2,494	2,425	2,352	2,276
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494	2,450	2,381	2,308	2,230
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456	2,412	2,342	2,269	2,191
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423	2,378	2,308	2,234	2,155
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348	2,278	2,203	2,124

Número de graus de liberdade do numerador.

Número de graus de liberdade do denominador.

Microsoft Excel
Microsoft Office XP
=INVF(0,05;7;15) = 2,707

12

1. Análise de variâncias

1.1 ANOVA de um factor

Avaliação da origem da diferença observadas na ANOVA (1/1):

A diferença entre as duas estimativas de σ pode ter diversas origens: por exemplo, uma média pode diferir das outras, todas as médias podem ser diferentes entre si, podem existir grupos de médias equivalentes, etc.

Uma forma simples de avaliar a razão desta diferença é através da ordenação das médias por ordem crescente e comparação das diferenças entre valores adjacentes com uma grandeza designada como “menor diferença significativa”, mds:

$$\circ \quad mds = s_{intra-a} \sqrt{(2/n)} \times t_{[h(n-1); 0,05; bilateral]}$$

Em que $s_{intra-a}$ é a estimativa de σ baseada na variação “intra-amostra”:

$$\circ \quad s_{intra-a} = \sqrt{\sum_i \sum_j (x_{ij} - \bar{x}_i)^2 / h(n-1)}$$

Importante: Este teste é aproximado!

13

1. Análise de variâncias

1.2 A aritmética da ANOVA

Se H_0 for verdadeira pode estimar-se σ de outra forma:

Se H_0 for verdadeira, pode estimar-se σ^2 tratando os dados como uma grande amostra. Este cálculo envolve o cálculo do somatório dos desvios em relação à média global \bar{x} :

$$\sum_i \sum_j (x_{ij} - \bar{x})^2$$

...e a divisão deste somatório pelo número de graus de liberdade ($hn-1$).

$$\sigma^2 = \frac{\sum_i \sum_j (x_{ij} - \bar{x})^2}{nh-1}$$

Importante: Este termo não é usado nos cálculos da ANOVA porque é função de ambas as variabilidades “intra” e “inter-amostra”.

14

1. Análise de variâncias

1.2 A aritmética da ANOVA

Existe uma relação algébrica exacta entre esta variação total e as fontes de variação que contribuem para si (ver Tabela):

Fontes de variação	Somas dos quadrados	Graus de liberdade
Intre-amostras	$n \sum_i (\bar{x}_i - \bar{x})^2$	$h - 1$
Intra-amostras	$\sum_i \sum_j (x_{ij} - \bar{x}_i)^2$	$h(n - 1)$
Total	$\sum_i \sum_j (x_{ij} - \bar{x})^2$	$hn - 1$

Os valores da última linha da tabela resultam da soma dos valores da duas primeiras linhas, tanto para a soma dos quadrados como para o número de graus de liberdade.

15

1. Análise de variâncias

1.2 A aritmética da ANOVA

Existe uma relação algébrica exacta entre esta variação total e as fontes de variação que contribuem para si (ver Tabela):

Fontes de variação	Somas dos quadrados	Graus de liberdade
Intre-amostras	$n \sum_i (\bar{x}_i - \bar{x})^2$	$h - 1$
Intra-amostras	$\sum_i \sum_j (x_{ij} - \bar{x}_i)^2$	$h(n - 1)$
Total	$\sum_i \sum_j (x_{ij} - \bar{x})^2$	$hn - 1$

Os valores da última linha da tabela resultam da soma dos valores da duas primeiras linhas, tanto para a soma dos quadrados como para o número de graus de liberdade.

16

1. Análise de variâncias

1.3. Avaliação da incerteza da amostragem

A incerteza padrão associada à amostragem, u_S , dum sistema heterogéneo pode ser avaliada através de uma ANOVA de um factor: “heterogeneidade do sistema” vs. “repetibilidade da medição”:

Incerteza de uma amostragem única, u_S :

$$u_S = s_{inter-am} = \sqrt{\frac{n \sum_{i=1}^n (\bar{x}_i - \bar{x})^2}{h - 1}}$$

Incerteza de uma amostragem composta de m amostras, $u_{S,m}$:

$$u_{S,m} = \frac{u_S}{\sqrt{m}}$$

Incerteza global expandida, U (n.c. aprox. 95 %):

$$U = 2 \sqrt{u_{S,m}^2 + u_{AA}^2}$$

↖ Análise da amostra

17

1. Análise de variâncias

1.4. Avaliação de tendências

Avaliação da relevância de variações da contaminação ambiental observada em condições equivalentes (tipicamente condições atmosféricas).

Se os resultados em dois dias, 1 e 2, forem:

Dia 1: $(C_1 \pm 2u_1)$ unidades

Dia 2: $(C_2 \pm 2u_2)$ unidades

em que u_1 e u_2 representam as incerteza padrão das medições de C_1 e C_2 .

A variação da concentração é relevante se:

$$|C_1 - C_2| > 2 \sqrt{u_1^2 + u_2^2}$$

A avaliação é realizada para um nível de confiança de aproximadamente 95%.

18

1. Análise de variâncias

1.4. Exercício

Exercício:

Estudo da incerteza da amostragem: Considere as seguintes estimativas de NO_x (em $\mu\text{mol L}^{-1}$) em quatro pontos, equidistantes, que definem uma linha que atravessa toda a largura de um rio que flui de Este para Oeste.

	Pontos de amostragem	Medições Replicadas	Média
A	Ponto mais a norte	102; 100; 101	101
B	Segundo ponto mais a norte	101; 101; 104	102
C	Segundo ponto mais a sul	97; 95; 99	97
D	Ponto mais a sul	90; 92; 94	92
			98

Avalie a incerteza da amostragem e da análise de uma amostras composta de duas amostras em que a análise de amostras tem uma incerteza padrão relativa de 2.2 %.

19

1. Análise de variâncias

Quando se utiliza uma ANOVA de dois factores? (1/3)

Quando se estuda o efeito de dois factores num determinado resultado?

Os resultados são apresentados numa tabela equivalente à usada na ANOVA de um factor, o entanto, os dados estão ordenados considerando as linhas e as colunas (i.e., são classificado de acordo com dois factores).

Exemplo: Comparação do rendimento da extracção de iões metálicos de uma solução aquosa considerando diferentes agentes quelantes.

○	Agente Quelante			
	Dia	A	B	C
1	84	80	83	79
2	79	77	80	79
3	83	78	80	78

20

1. Análise de variâncias

Quando se utiliza uma ANOVA de dois factores? (2/3)

Exemplo: Comparação do rendimento da extracção de iões metálicos de uma solução aquosa considerando diferentes agentes quelantes.

Dia	Agente Quelante			
	A	B	C	D
1	84	80	83	79
2			80	79
3			80	78

Mesmo dia =
Mesma Solução &
Condições Ambientais

Em cada dia, foi preparada uma solução fresca de metal (com uma determinada concentração) e procedeu-se a extracção do metal considerando cada um dos agentes quelantes, em que as extracções foram realizadas por ordem aleatória.

21

1. Análise de variâncias

Quando se utiliza uma ANOVA de dois factores? (3/3)

Exemplo: Comparação do rendimento da extracção de iões metálicos de uma solução aquosa considerando diferentes agentes quelantes.

Dia	Agente Quelante			
	A	B	C	D
1	84	80	83	79
2	79	77	80	79
3	83	78	80	78

A ANOVA permite avaliar se:

1) as eficiência de reacção associadas aos diferentes agentes quelantes são significativamente diferentes;

OU

2) a variação de dia para dia é significativamente superior à variação devida aos erros aleatórios da medição.

22

Realização de testes em Excel

1

2

3

4

5

- Activação da Análise de Dados:
 - ▶ Ferramentas → Suplementos → (...)
 - (...) → “ Analysis Toolpak”.
- Utilização da Análise de Dados:
 - ▶ Ferramentas → Análise da dados → (...).

23