

Introdução à análise de dados longitudinais

Prof. Caio Azevedo

Introdução

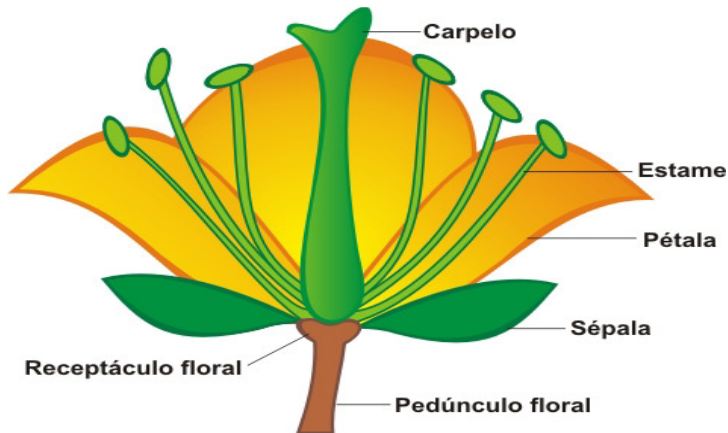
- (Dados de) Medidas repetidas: medidas feitas nas mesmas unidades experimentais ao longo de alguma condição de avaliação (distância, peso, tempo etc).
- Dados longitudinais: medidas repetidas nas quais os níveis das condições de avaliação (ou simplesmente “condição”) não podem ser aleatorizados entre si (por exemplo, o tempo ou profundidade).
 - Univaridos: uma única variável medida ao longo das condições.
 - Multivariados: mais de uma variável medida ao longo das condições.

- Diferença entre dados longitudinais, dados multivariados e séries temporais.
 - Dados longitudinais: uma ou mais de uma variável medida de muitos indivíduos ao longo de (poucas) condições (mesmo sendo o tempo, profundidade).
 - Séries temporais: uma ou mais de uma variável medida de poucos (ou somente um) indivíduo(s) (que também podem fazer papel de variável) ao longo de muitos instantes no tempo.
 - Dados multivariados: mais de uma variável (de diferentes naturezas) medidas, em geral, em um único instante (cohort).
- No entanto, pode-se tomar emprestado técnicas desenvolvidas para análise de dados multivariados e/ou de séries temporais para a análise de dados longitudinais.

Exemplo de dados multivariados: Iris de Fisher

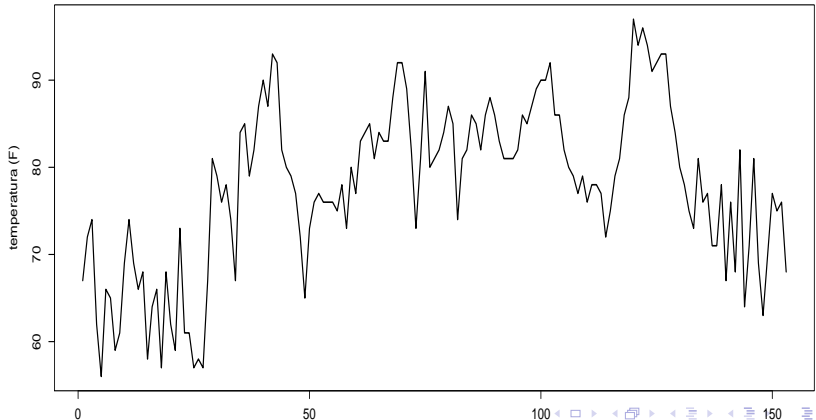
- Os dados consistem de 50 unidades amostrais de três espécies (setosa, virginica, versicolor) de íris (uma espécie de planta), ou seja, temos um total de 150 unidades amostrais.
- De cada uma delas mediu-se quatro variáveis: comprimento e largura da sépala (CS, LS) e comprimento e largura da pétala (CP,LP).
- Objetivo original: quantificar a variação morfológica em relação à essas espécies com bases nas quatro variáveis de interesse.

Cont.



Planta	Tipo	Comp. Sep	Larg. Sep.	Comp. Pet.	Larg. Pet
1	setosa	5,10	3,50	1,40	0,20
2	setosa	4,90	3,00	1,40	0,20
⋮	⋮	⋮	⋮	⋮	⋮
50	setosa	5,00	3,30	1,40	0,20
51	versicolor	7,00	3,20	4,70	1,40
52	versicolor	6,40	3,20	4,50	1,50
⋮	⋮	⋮	⋮	⋮	⋮
100	versicolor	5,70	2,80	4,10	1,30
101	virginica	6,30	3,30	6,00	2,50
102	virginica	5,80	2,70	5,10	1,90
⋮	⋮	⋮	⋮	⋮	⋮
150	virginica	5,90	3,00	5,10	1,80

Exemplo de séries temporais: temperatura em ($^{\circ}$ F) diária de Nova York (Maio a Setembro de 1973)



Exemplo de dados longitudinais: Estudo da eficácia de escovas de dentes

- Considere o seguinte estudo na área de Odontopediatria.
- O objetivo é comparar duas escovas de dente (convencional e experimental, chamada de “hugger”) com respeito à redução de um índice de placa bacteriana (IPB) em crianças de ambos os sexos em idade pré-escolar. Os valores obtidos correspondem aos IPB’s medidos em alguns dentes antes e depois da escovação dental de 14 crianças do sexo feminino e 12 do sexo masculino. Cada criança utilizou cada um dos tipos de escova sendo sempre a experimental, a primeira. O tipo de escova tende a ser melhor quanto maior for sua “capacidade de remoção” da placa bacteriana.

Dados

Criança	Tipo de escova				
	Sexo	Hugger		Convencional	
		Antes	Depois	Antes	Depois
1	F	2,18	0,43	1,2	0,75
2	F	2,05	0,08	1,43	0,55
⋮	⋮	⋮	⋮	⋮	⋮
25	M	1,3	0,05	2,73	0,85
26	M	2,65	0,25	3,43	0,88

Outro estudo de eficácia de escovas de dentes

- Considere o estudo na área de Odontopediatria realizado na Faculdade de Odontologia da Universidade de São Paulo por Célia Regina Rodrigues e Symonne Parizotto.
- O objetivo é comparar duas escovas de dente (convencional e monobloco) com respeito à redução de um índice de placa bacteriana (IPB).
- Os valores obtidos correspondem a índices de placa bacteriana medidos nos dentes posteriores (pré-molares e molares) antes e depois da escovação dental de 32 crianças entre 4 e 6 anos de idade. O tipo de escova tende a ser melhor quanto maior for sua “capacidade de remoção” da placa bacteriana.

Dados

Tipo de escova	Sessão	Antes	Depois	Indivíduo
CT	1	1,05	1,00	1
CT	2	1,13	0,84	1
CT	3	1,15	0,86	1
CT	4	1,13	0,94	1
⋮	⋮	⋮	⋮	⋮
CT	1	1,4	1,12	16
CT	2	1,25	0,67	16
CT	3	1,5	1,1	16
CT	4	1,5	1,22	16
MT	1	1,66	1,63	17
MT	2	1,36	1,16	17
MT	3	1,52	0,88	17
MT	4	1,41	1,20	17
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
MT	1	1,15	1	32
MT	2	1,23	1,11	32
MT	3	1,15	1,07	32
MT	4	1,26	1	32

Exemplo de dados longitudinais: efeitos de agentes oxidantes no sistema respiratório

- Estudo realizado na Faculdade de Medicina da Universidade de São Paulo para avaliar os efeitos de agentes oxidantes no sistema respiratório.
- Espera-se que a exposição à maiores concentrações desses agentes causem danos crescentes aos cílios e às células secretoras de muco, os principais meios de defesa do sistema respiratório.
- 56 palatos de rãs foram aleatoriamente alocados a seis grupos com 10 palatos em cinco deles e 6 em outro.

Cont.

- Os palatos de cada um desses grupos foram imersos por 35 minutos numa solução de peróxido de hidrogênio com uma concentração especificada, a saber, 0, 1, 8, 16, 32 ou 64 microM (que correspondem aos grupos).
- A velocidade de transporte mucociliar (mm/s) (variável resposta) foi observada a cada 5 minutos após a imersão (quanto maior, melhor).

(microM)	5	10	15	20	25	30	35
0	1,24	1,21	1,16	1,08	1,08	0,87	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	0,84	0,89	0,81	0,84	1,05	0,95	
1	1,12	1,22	1,14	1,22	1,46	1,76	1,72
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	0,89	0,84	0,93	1,02	0,93	0,89	0,86
8	0,81	1,18	0,99	1,08	1,00	1,09	1,09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8	1,26	1,24	1,20	1,13	1,03	1,08	1,13
16	0,88	0,73	0,61	0,58	0,61	0,48	0,52
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	1,08	1,12	0,86	1,02	0,85	0,90	0,89
32	0,97	0,67	0,49	0,46	0,38	0,35	0,29
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
32	0,72	0,48	0,63	0,57	0,63	0,53	0,54
64	0,12	0,09	0,22	0,11	0,16	0,11	0,09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
64	0,23	0,40	0,25	0,16	0,27	0,27	0,47

Exemplo de dados longitudinais: impacto da radiação de “laser” na microdureza média do esmalte dentário de dentes expostos a desafios cariogênicos

- Três blocos de esmalte dentário foram confeccionados a partir de cada um de 32 terceiros molares humanos extraídos por indicação ortodôntica, totalizando 96 blocos.
- Esses blocos, após limpos, foram examinados para verificação de defeitos e blocos com defeitos foram substituídos.

Cont.

- Em seguida, os blocos sem defeitos foram limpos e subsequentemente, alocados aleatoriamente a um de três tratamentos de forma balanceada (32 blocos por tratamento), a saber: Controle extra bucal, o Controle in situ e Laser.
- Após certos procedimentos medidas da microdureza em diferentes profundidades (10, 20, 40, 60, 120 e 180 μm) em cada bloco foram observadas em triplicata.

Existem mais duas tabelas semelhantes à essa

Voluntário	Dente	Profundidade μm						
		10	20	40	60	80	120	180
1	1	162,33	135,67	166,67	207,00	290,00	289,67	266,67
1	2	193,67	168,33	178,00	196,33	275,00	317,33	330,67
1	3	159,33	165,33	165,33	255,67	295,33	308,67	319,00
1	4	189,00	175,67	162,00	218,00	293,67	273,33	287,33
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8	29	167,33	171,67	205,33	261,67	296,33	294,33	296,33
8	30	159,00	159,33	197,00	246,33	277,33	302,67	303,33
8	31	135,33	173,00	158,67	250,67	292,67	289,33	300,33
8	32	170,33	153,67	148,67	243,33	302,33	285,33	286,33

Exemplo de dados longitudinais: estudo do diâmetro diastólico da aorta em recém-nascidos pré-termo

- Um dos objetivos era estimar a curva média de crescimento do diâmetro diastólico da aorta por unidade de peso (AO) em mm/kg para recém-nascidos pré-termo pequenos para a idade gestacional (PIG) no período compreendido entre a 26^a à 40^a semanas pós concepção.

RNPT	Semanas após o nascimento														
	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1	7.3	9.4	10.3		8.7	8.2	6.7	6.1		5.6		5.1			4.9
			6.1	6.1	6.2		5.4	5.2	4.9						
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
60		8.5													
61				10.9	10.7	9.4		8			5.8		4.9		

Comentários

- Espera-se observar dependência entre as medidas feitas nas mesmas unidades experimentais.
- Não modelar a estrutura de dependência ou modelá-la de forma inadequada pode levar à uma subestimação dos erros-padrão de alguns dos (ou todos) parâmetros de interesse.

Objetivos gerais

- Apresentar ferramentas de estatística descritivas apropriadas.
- Modelar a estrutura de dependência de forma apropriada (descritivo e inferencial).
- Modelagem: modelos paramétricos em que a estrutura de dependência será traduzida em termos probabilísticos e/ou em termos de uma estrutura de correlação.

Modelagem da estrutura de dependência

- Toda metodologia de análise multivariada busca estudar e/ou compreender e/ou utilizar, quando existem, estruturas de dependência e/ou correlação.
 - Estrutura de dependência: modelo probabilístico ou modelo estatístico (modelo de regressão), ou seja, a distribuição conjunta das variáveis (resposta) de interesse.
 - Estrutura de correlação: Pearson (variáveis contínuas e/ou discretas com suporte correspondendo à um conjunto infinito), Spearman, Kendall (não paramétricas), policórica, tetracórica (variáveis discretas com suporte correspondendo à um conjunto finito).

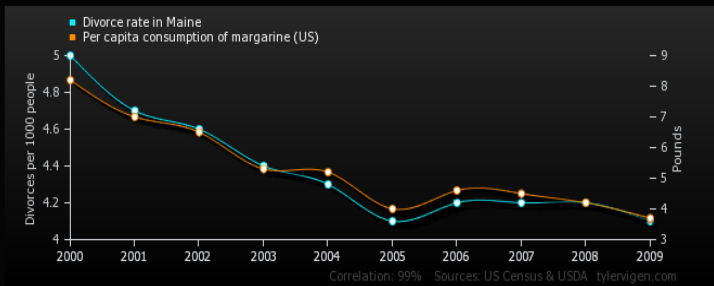
Correlação

- Existência de correlação (correlação significativa) não implica numa relação de causalidade (do ponto de vista do problema).
 - Em geral, altura e peso são positivamente correlacionados, mas a altura não é determinada (biologicamente) pelo peso e vice-versa.
 - Outros fatores: genética, alimentação, meio-ambiente, de fato determinam (simultaneamente) a altura e peso.
- Os dois gráficos a seguir foram extraídos do site

http:

`//www.fastcodesign.com/3030529/infographic-of-the-day/
hilarious-graphs-prove-that-correlation-isnt-causation`

Número de divórcios em Maine × Consumo per capita de margarina (EUA)



Número de pessoas que se afogaram em piscinas × número de filmes em que o Nicolas Cage apareceu

