

Analiza wariancji dla wielowymiarowych danych funkcjonalnych

Łukasz Smaga

Wydział Matematyki i Informatyki
Uniwersytet im. Adama Mickiewicza w Poznaniu

Cykl wykładów otwartych „Wielowymiarowe metody statystyczne”

- Niech

$$\mathbf{X}_{ij}(t) = (X_{ij1}(t), X_{ij2}(t), \dots, X_{ijp}(t))^{\top}, \quad i = 1, 2, \dots, l, \quad j = 1, 2, \dots, n_i$$

oznaczają l grup wektorowych procesów losowych zdefiniowanych na $T = [a, b]$.

- Przez $SP_p(\mathbf{m}, \Gamma)$ oznaczamy p -wymiarowy proces stochastyczny o średniej $\mathbf{m}(t)$, $t \in T$ i funkcji kowariancji $\Gamma(s, t)$, $s, t \in T$.
- Załóżmy, że \mathbf{X}_{ij} są niezależnymi procesami stochastycznymi $SP_p(\mathbf{m}_i, \Gamma)$, $i = 1, 2, \dots, l$, $j = 1, 2, \dots, n_i$ i rozważmy następujący układ hipotez

$$H_0 : \mathbf{m}_1(t) = \mathbf{m}_2(t) = \dots = \mathbf{m}_l(t), \quad t \in T, \quad H_1 : \neg H_0. \quad (1)$$

- Załóżmy, że $\mathbf{X}_{ij} \in L_2^p(T)$, $i = 1, 2, \dots, l$, $j = 1, 2, \dots, n_i$, gdzie $L_2^p(T)$ jest przestrzenią Hilberta p -wymiarowych wektorów funkcji całkowlanych z kwadratem na przedziale T z iloczynem skalarnym $\langle \mathbf{x}, \mathbf{y} \rangle = \int_T \mathbf{x}^\top(t) \mathbf{y}(t) dt$.
- Przyjmujemy, że

$$X_{ijm}(t) = \sum_{r=0}^{K_m} \alpha_{ijmr} \varphi_r(t), \quad t \in T, \quad m = 1, 2, \dots, p, \quad (2)$$

gdzie α_{ijmr} , $r = 0, 1, \dots, K_m$ są zmiennymi losowymi o skończonej wariancji oraz $\{\varphi_r\}$ jest bazą w $L_2(T)$.

- Niech $KM = \max\{K_1, K_2, \dots, K_p\}$, $\alpha_{ijm} = (\alpha_{ijm0}, \alpha_{ijm1}, \dots, \alpha_{ijmK_m}, 0, \dots, 0) \in \mathbb{R}^{KM+1}$ oraz $\varphi(t) = (\varphi_0(t), \varphi_1(t), \dots, \varphi_{KM}(t))^T$, $t \in T$. Wtedy

$$\mathbf{X}_{ij}(t) = \begin{pmatrix} \alpha_{ij1} \\ \alpha_{ij2} \\ \vdots \\ \alpha_{ijp} \end{pmatrix} \varphi(t) = \alpha_{ij} \varphi(t), \quad t \in T. \quad (3)$$

- Osobno dla każdej zmiennej, wektory $\alpha_{ijm}^* = (\alpha_{ijm0}, \alpha_{ijm1}, \dots, \alpha_{ijmK_m})^T$, $i = 1, 2, \dots, l$, $j = 1, 2, \dots, n_i$ mogą być estymowane za pomocą metody najmniejszych kwadratów, a optymalna wartość K_m może być wybrana poprzez Bayesowskie kryterium informacyjne (BIC), rozszerzone Bayesowskie kryterium informacyjne (eBIC) lub kryterium informacyjne Akaike (AIC). Wówczas spośród wartości K_m odpowiadających wszystkim procesom wybieramy wartość modalną, minimalną, maksymalną lub średnią jako wspólną wartość dla wszystkich $(X_{ijm}(t), t \in T)$, $i = 1, 2, \dots, l$, $j = 1, 2, \dots, n_i$.

- Podobnie do testów dla MANOVA (Anderson, 2003), poniższe macierze zostaną wykorzystane do konstrukcji statystyk testowych dla FMANOVA:

$$\mathbf{E} = \sum_{i=1}^l \sum_{j=1}^{n_i} \int_T (\mathbf{X}_{ij}(t) - \bar{\mathbf{X}}_i(t)) (\mathbf{X}_{ij}(t) - \bar{\mathbf{X}}_i(t))^{\top} dt,$$

$$\mathbf{H} = \sum_{i=1}^l n_i \int_T (\bar{\mathbf{X}}_i(t) - \bar{\mathbf{X}}(t)) (\bar{\mathbf{X}}_i(t) - \bar{\mathbf{X}}(t))^{\top} dt,$$

gdzie

$$\bar{\mathbf{X}}_i(t) = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{X}_{ij}(t) = \frac{1}{n_i} \left(\sum_{j=1}^{n_i} \alpha_{ij} \right) \varphi(t), \text{ dla } i = 1, 2, \dots, l,$$

$$\bar{\mathbf{X}}(t) = \frac{1}{n} \sum_{i=1}^l \sum_{j=1}^{n_i} \mathbf{X}_{ij}(t) = \frac{1}{n} \left(\sum_{i=1}^l \sum_{j=1}^{n_i} \alpha_{ij} \right) \varphi(t), \text{ } t \in T, \text{ } n = n_1 + n_2 + \dots + n_l.$$

Twierdzenie 1

Jeżeli składowe procesów stochastycznych $(\mathbf{X}_{ij}(t), t \in T)$, $i = 1, 2, \dots, l$, $j = 1, 2, \dots, n_i$, mogą być reprezentowane przez skończoną liczbę funkcji bazowych jak w (3), to $\mathbf{E} = \mathbf{A} - \mathbf{B}$ oraz $\mathbf{H} = \mathbf{B} - \mathbf{C}$, gdzie

$$\mathbf{A} = \sum_{i=1}^l \sum_{j=1}^{n_i} \alpha_{ij} \mathbf{J}_\varphi \alpha_{ij}^\top, \quad \mathbf{B} = \sum_{i=1}^l \frac{1}{n_i} \sum_{j=1}^{n_i} \sum_{m=1}^{n_i} \alpha_{ij} \mathbf{J}_\varphi \alpha_{im}^\top, \quad \mathbf{C} = \frac{1}{n} \sum_{i=1}^l \sum_{j=1}^{n_i} \sum_{t=1}^l \sum_{u=1}^{n_t} \alpha_{ij} \mathbf{J}_\varphi \alpha_{tu}^\top,$$

gdzie $\mathbf{J}_\varphi := \int_T \varphi(t) \varphi^\top(t) dt$ jest macierzą iloczynów (ang. cross product matrix) odpowiadającą φ wymiaru $KM \times KM$.

- Rozważamy następujące statystyki testowe, które są skonstruowane podobnie do statystyk dla MANOVA (Anderson, 2003):
 - 1 $W = \det(\mathbf{E}) / \det(\mathbf{E} + \mathbf{H})$ – statystyka Λ Wilksa
 - 2 $LH = \text{trace}(\mathbf{H}\mathbf{E}^{-1})$ – statystyka T^2 Lawleya-Hotellinga
 - 3 $P = \text{trace}(\mathbf{H}(\mathbf{H} + \mathbf{E})^{-1})$ – statystyka Pillaia
 - 4 $R = \lambda_{\max}(\mathbf{H}\mathbf{E}^{-1})$ – statystyka Roya
- Hipotezę zerową odrzucamy przy małych wartościach statystyki W oraz dużych wartościach statystyk LH , P i R .
- Przybliżenia rozkładu statystyk testowych przy prawdziwości hipotezy zerowej dokonamy za pomocą metody permutacyjnej. Testy permutacyjne mogą być jednak czasochłonne.

Uwaga 1

Sumy \mathbf{A} i \mathbf{C} są niezmiennicze względem permutacji danych $(\mathbf{X}_{ij}(t), t \in T)$, $i = 1, 2, \dots, l$, $j = 1, 2, \dots, n_j$.

- Zatem tylko suma \mathbf{B} musi być policzona dla każdej losowej permutacji danych z osobna.

- Drugą grupą testów są procedury oparte o losowe projekcje danych funkcjonalnych na dane wektorowe. Idea takich testów pochodzi od Cuesta-Albertos i Febrero-Bande (2010).
- Załóżmy, że \mathcal{H} jest separowalną przestrzenią Hilberta wyposażoną w iloczyn skalarny $\langle \cdot, \cdot \rangle$, a μ jest rozkładem gaussowskim na tej przestrzeni takim, że każda jego jednowymiarowa projekcja jest niezdegenerowana. Ponadto, niech $\mathbf{m}_i = (m_{i1}, m_{i2}, \dots, m_{ip})^\top$, $m_{ij} \in \mathcal{H}$, $i = 1, 2, \dots, l$, $j = 1, 2, \dots, p$, oraz $v_m \in \mathcal{H}$ będzie elementem losowym wybranym zgodnie z rozkładem μ , $m = 1, 2, \dots, p$. Prawdziwość hipotezy $H_0 : \mathbf{m}_1(t) = \mathbf{m}_2(t) = \dots = \mathbf{m}_l(t)$, $t \in T$ implikuje prawdziwość poniższej hipotezy zerowej:

$$H_0^{\mathbf{V}} : \begin{pmatrix} \langle v_1, m_{11} \rangle \\ \vdots \\ \langle v_p, m_{1p} \rangle \end{pmatrix} = \begin{pmatrix} \langle v_1, m_{21} \rangle \\ \vdots \\ \langle v_p, m_{2p} \rangle \end{pmatrix} = \dots = \begin{pmatrix} \langle v_1, m_{l1} \rangle \\ \vdots \\ \langle v_p, m_{lp} \rangle \end{pmatrix}$$

dla każdego $\mathbf{V} = (v_1, v_2, \dots, v_p)^\top \in \mathcal{H} \times \dots \times \mathcal{H}$.

- Jednak co można powiedzieć o $H_0^{\mathbf{V}}$, gdy H_0 nie jest prawdziwa?

Twierdzenie 2

Przy powyższych założeniach odnośnie \mathcal{H} oraz μ , jeżeli $m_{ij} \in \mathcal{H}$, $i = 1, 2, \dots, l$, $j = 1, 2, \dots, p$, oraz istnieją r_1, r_2, s takie, że $m_{r_1 s} \neq m_{r_2 s}$, to $(\mu \times \dots \times \mu)(\mathcal{A}) = 0$, gdzie $\mu \times \dots \times \mu$ jest miarą produktową na przestrzeni $\mathcal{H} \times \dots \times \mathcal{H}$ oraz zbiór \mathcal{A} zawiera wszystkie wektory $\mathbf{V} = (v_1, v_2, \dots, v_p)^\top \in \mathcal{H} \times \dots \times \mathcal{H}$, dla których prawdziwa jest hipoteza $H_0^{\mathbf{V}}$.

- Twierdzenie 2 implikuje, że jeżeli hipoteza H_0 nie jest prawdziwa, to dla $(\mu \times \dots \times \mu)$ -prawie każdego $\mathbf{V} \in \mathcal{H} \times \dots \times \mathcal{H}$, hipoteza $H_0^{\mathbf{V}}$ również nie jest prawdziwa.
- Podsumowując, wydaje się, że test statystyczny na poziomie istotności $\alpha \in (0, 1)$ dla $H_0^{\mathbf{V}}$ powinien być również testem statystycznym na zbliżonym poziomie istotności dla H_0 . Innymi słowy, test dla MANOVA może być wykorzystany do FMANOVA, poprzez zastosowanie go do testowania $H_0^{\mathbf{V}}$.

- Załóżmy, że $\mathbf{X}_{ij} \in L_2^p(T)$, $i = 1, 2, \dots, l$, $j = 1, 2, \dots, n_i$. Procedura testowa jest następująca:
- ① Wybierz, zgodnie z rozkładem gaussowskim, funkcje v_m , $m = 1, 2, \dots, p$ w przestrzeni $L_2(T)$.
- ② Wyznacz projekcje

$$Y_{ijm} = \frac{\int_T X_{ijm}(t)v_m(t)dt}{(\int_T v_m^2(t)dt)^{1/2}}$$

dla $i = 1, 2, \dots, l$, $j = 1, 2, \dots, n_i$, $m = 1, 2, \dots, p$.

- ③ Zastosuj test MANOVA do l prób wektorów losowych $\mathbf{Y}_{ij} = (Y_{ij1}, Y_{ij2}, \dots, Y_{ijp})^\top$, $i = 1, 2, \dots, l$, $j = 1, 2, \dots, n_i$.
- Aby wykonać krok 1, można wykorzystać biały szum gaussowski lub ruchy Browna.
- W kroku 3 procedury, należy wybrać odpowiedni test MANOVA. Wykorzystamy standardowe testy MANOVA: test Λ Wilksa (test Wp), test T^2 Lawleya-Hotellinga (test LHp), test Pillaia (test Pp), oraz test Roya (test Rp) lub ich wersje permutacyjne.

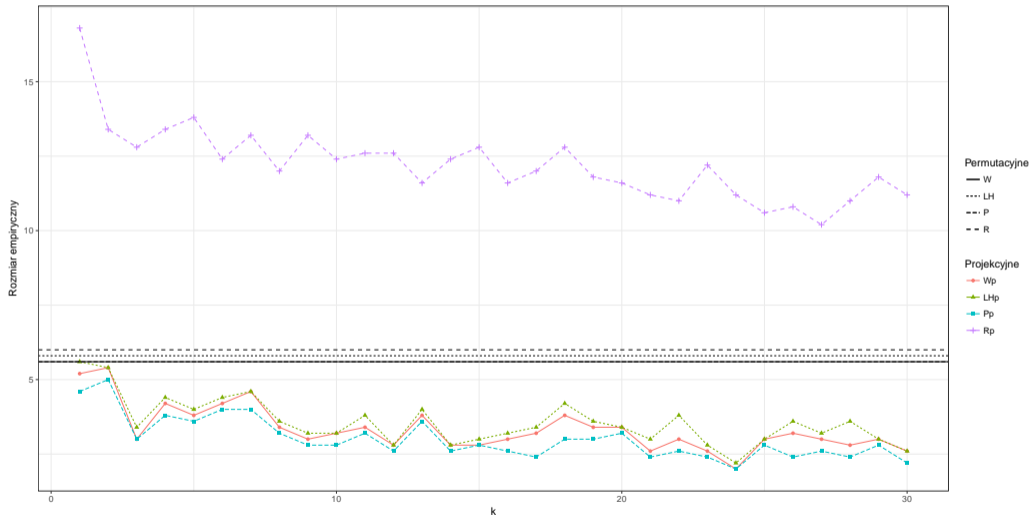
- Powyższa procedura ma dwie wady.
- Zastępując funkcję przez tylko jedną liczbę rzeczywistą, tracimy część informacji. To może spowodować utratę mocy.
- Stosując losowo wybraną projekcję, akceptujemy pewną losową niestabilność w procedurze testowej, ponieważ powtarzając procedurę dwukrotnie, hipoteza zerowa może zostać odrzucona za pierwszym razem, a za drugim możemy nie mieć podstaw do jej odrzucenia.

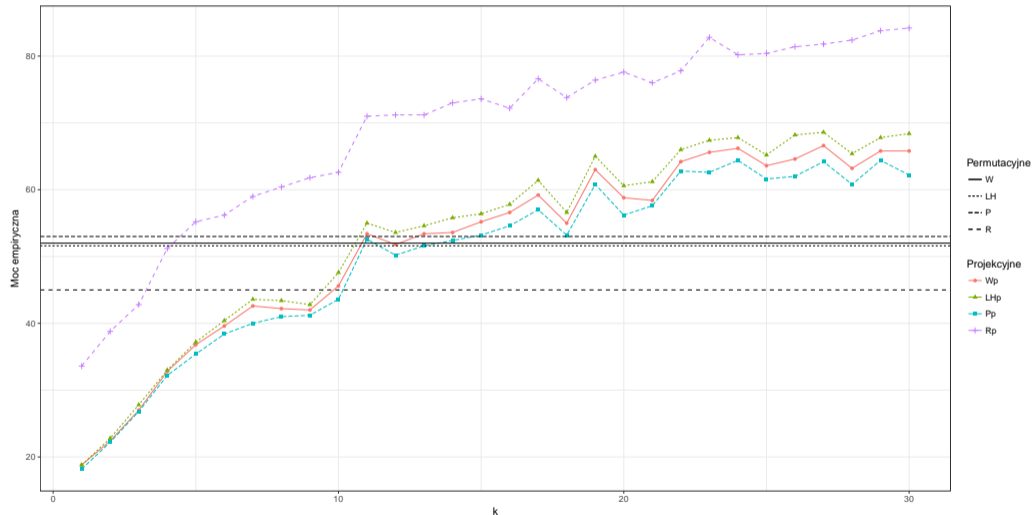
- Aby uniknąć takich problemów, powyższą procedurę wykonuje się wielokrotnie dla $k > 1$ projekcji, a następnie koryguje się p -wartości tak aby kontrolować frakcję fałszywych odkryć (FDR, ang. False Discovery Rate). Korzystając z wyników Benjamini i Yekutieli (2001), p -wartość ma postać

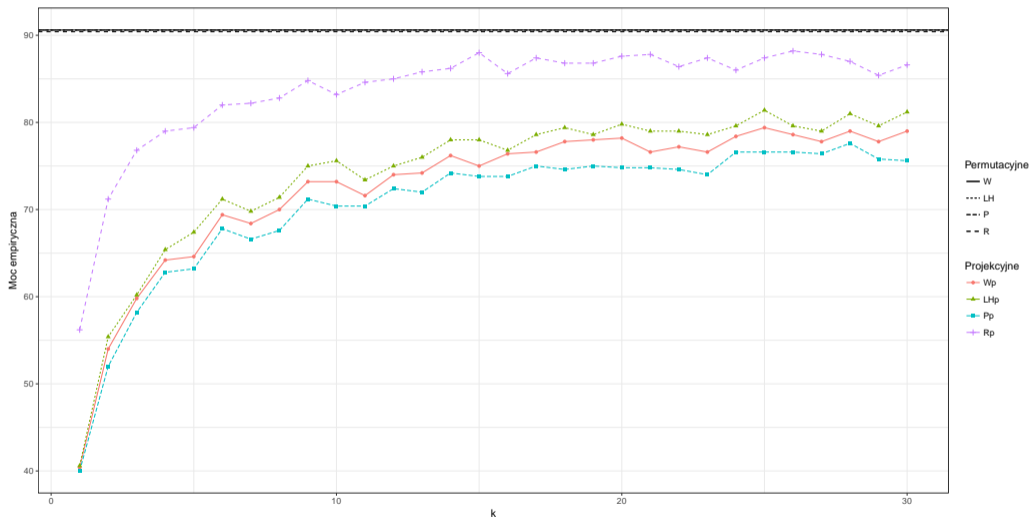
$$\inf \left\{ \frac{kp_{(i)}}{i}, i = 1, 2, \dots, k \right\},$$

gdzie $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k)}$ są uporządkowanymi p -wartościami otrzymanymi w k testach dla tej samej hipotezy zerowej.

- Rekomendowana liczba projekcji k jest liczbą bliską 30, ale w przypadkach wątpliwych (np., gdy p -wartość jest bliska poziomowi istotności), należy użyć większej liczby projekcji k .







- Rozmiary empiryczne ($\alpha = 5\%$) testów opartych o losowe projekcje ($k = 30$) dla różnych metod generowania procesów gaussowskich („s” - standardowe testy MANOVA na projekcjach, „p” - testy permutacyjne na projekcjach). Im większe ρ tym mniejsza korelacja.

Rozkład	n	ρ	V	Gauss. biały szum				Ruchy Browna			
				Wp	LHp	Pp	Rp	Wp	LHp	Pp	Rp
$t_4/\sqrt{2}$	(10, 20, 15)	0,1	s	3,1	3,6	2,3	11,6	3,2	3,5	2,8	11,6
			p	3,7	3,9	3,9	4,0	3,7	3,8	3,7	3,7
		0,3	s	2,7	4,1	2,0	12,5	3,2	3,8	2,8	10,5
			p	4,0	4,4	4,4	4,4	3,7	4,0	3,9	4,0
		0,5	s	3,3	4,5	2,4	14,8	3,4	4,1	2,8	10,8
			p	4,0	3,7	4,3	4,2	4,1	4,1	4,6	4,1
		0,7	s	4,5	5,6	3,2	16,9	3,4	4,5	2,5	11,9
			p	4,8	4,8	4,7	5,0	4,8	4,7	4,7	4,1
		0,9	s	5,0	6,1	3,3	18,2	4,2	4,8	3,3	13,1
			p	5,3	5,0	4,7	5,7	5,0	5,1	5,0	4,7

- Empiryczne moce ($\alpha = 5\%$) testów opartych o losowe projekcje ($k = 30$) dla różnych metod generowania procesów gaussowskich („s” - standardowe testy MANOVA na projekcjach, „p” - testy permutacyjne na projekcjach). Im większe ρ tym mniejsza korelacja.

Rozkład	n	ρ	δ	V	Gauss. biały szum				Ruchy Browna			
					Wp	LHp	Pp	Rp	Wp	LHp	Pp	Rp
$t_4/\sqrt{2}$	(10, 20, 15)	0,1	0,12	s	79,1	82,3	73,8	90,3	46,8	49,6	41,2	67,9
				p	81,0	81,7	80,2	82,3	48,0	48,8	46,5	49,4
		0,3	0,30	s	72,8	77,4	67,5	88,5	54,0	58,0	49,7	74,9
				p	76,5	76,5	75,6	76,8	56,7	57,2	56,5	56,4
		0,5	0,45	s	59,1	63,6	52,4	82,9	57,8	60,6	54,2	78,2
				p	63,3	64,3	62,3	66,3	61,1	61,0	60,1	59,5
		0,7	0,80	s	80,1	84,6	75,9	94,4	89,1	90,6	86,8	97,1
				p	83,1	83,4	82,5	83,4	90,4	90,4	90,2	90,5
		0,9	1,20	s	84,3	87,8	78,8	96,7	98,7	99,1	98,3	99,5
				p	86,5	86,6	85,7	86,9	98,8	98,8	98,4	98,7

- Empiryczne rozmiary ($\alpha = 5\%$) testów opartych o losowe projekcje ($k = 30$) przy równych i różnych funkcjach kowariancji w grupach. Dla testów permutacyjnych opartych o reprezentację bazową wyniki są podobne.

n_i	l	Równe funkcje kowariancji				Różne funkcje kowariancji			
		Wp	LHp	Pp	Rp	Wp	LHp	Pp	Rp
15	26	3,4	3,6	3,2	10,8	3,8	3,8	3,6	12,0
	81	3,4	4,0	3,2	11,6	4,0	4,4	3,6	9,6
	151	2,6	2,6	2,0	12,0	3,6	4,0	3,4	9,4
30	26	3,2	3,2	3,0	11,0	3,0	3,4	2,8	13,0
	81	4,4	4,4	4,0	11,2	2,2	2,4	2,2	9,6
	151	2,6	2,6	2,2	11,2	5,0	5,0	5,0	11,0
50	26	3,0	3,0	3,0	12,0	2,6	2,6	2,6	10,0
	81	4,2	4,4	4,0	11,4	3,8	3,8	3,4	12,6
	151	2,2	2,4	1,6	11,0	3,2	3,2	3,0	11,6

- Empiryczne moce ($\alpha = 5\%$) testów opartych o losowe projekcje ($k = 30$) przy równych i różnych funkcjach kowariancji w grupach. Dla testów permutacyjnych opartych o reprezentację bazową wyniki są podobne.

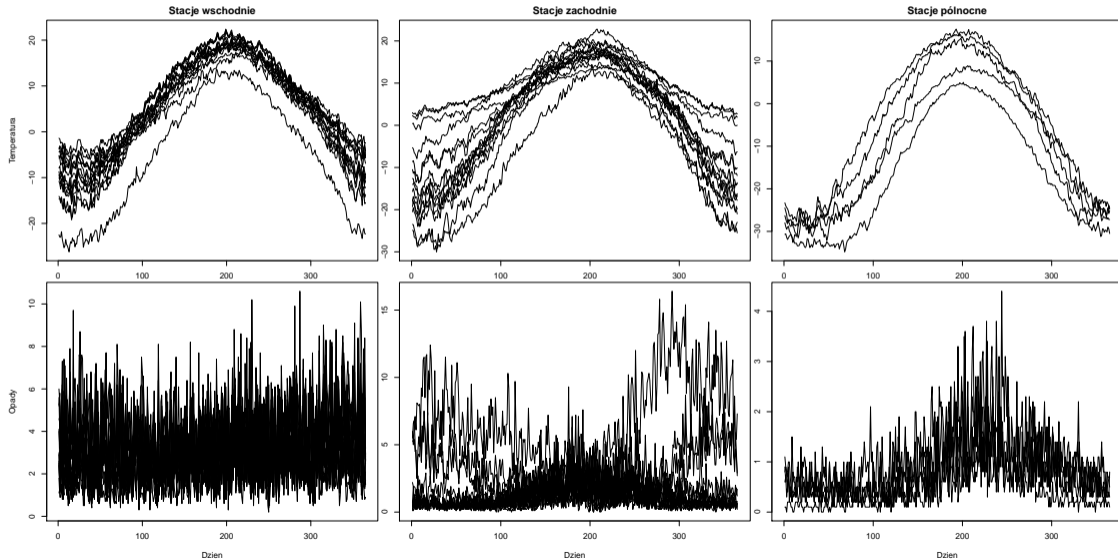
n_i	l	Równe funkcje kowariancji				Różne funkcje kowariancji			
		Wp	LHp	Pp	Rp	Wp	LHp	Pp	Rp
15	26	25,6	28,8	20,8	48,4	26,4	32,4	21,4	50,6
	81	28,0	31,6	23,8	47,8	27,0	30,4	20,6	48,8
	151	28,2	31,2	22,6	52,4	26,8	28,8	23,2	49,8
30	26	66,6	67,8	65,0	81,4	59,6	62,2	56,4	78,2
	81	67,0	68,2	66,6	82,4	70,4	72,8	68,6	86,8
	151	65,8	68,4	62,2	84,2	69,8	71,4	67,0	85,2
50	26	92,0	92,4	91,8	96,6	87,0	87,2	86,6	95,6
	81	89,0	89,8	87,8	95,6	91,6	92,0	91,2	96,8
	151	91,8	92,0	91,2	96,6	92,6	93,2	91,8	97,4

n_i	l	Permutacyjne				Projekcyjne ($k = 30$)				MANOVA średnie			
		W	LH	P	R	Wp	LHp	Pp	Rp	W	LH	P	R
15	26	15,6	15,0	15,4	12,8	25,6	28,8	20,8	48,4	5,0	5,6	4,8	15,8
	81	16,6	16,2	18,4	13,2	28,0	31,6	23,8	47,8	5,6	6,0	5,0	16,8
	151	19,8	19,2	20,4	15,4	28,2	31,2	22,6	52,4	6,0	5,8	5,6	17,6
30	26	43,0	43,8	43,4	36,8	66,6	67,8	65,0	81,4	7,0	7,0	6,8	18,2
	81	50,0	50,2	50,8	41,2	67,0	68,2	66,6	82,4	8,0	8,2	7,6	20,4
	151	52,0	51,6	53,0	45,0	65,8	68,4	62,2	84,2	7,6	8,0	7,6	20,4
50	26	87,8	88,2	87,6	87,4	92,0	92,4	91,8	96,6	7,8	7,8	7,6	21,4
	81	93,6	93,8	93,2	93,2	89,0	89,8	87,8	95,6	9,0	9,0	8,8	22,2
	151	93,8	93,8	94,0	94,0	91,8	92,0	91,2	96,6	7,6	7,6	7,6	21,4

n_i	l	Permutacyjne				Projekcyjne ($k = 30$)				MANOVA średnie			
		W	LH	P	R	Wp	LHp	Pp	Rp	W	LH	P	R
15	26	43,0	43,0	43,2	40,2	43,8	50,0	36,0	74,8	29,6	30,2	28,0	52,2
	81	44,4	44,2	44,4	42,2	47,8	53,4	42,2	74,0	30,2	31,4	29,6	54,6
	151	48,0	48,2	48,2	44,4	53,0	56,8	46,0	77,8	32,2	32,4	31,2	57,2
30	26	87,4	87,8	87,4	86,4	92,8	94,2	91,8	98,4	60,2	61,2	59,2	78,2
	81	86,6	86,8	86,6	87,2	91,6	92,4	91,4	97,2	65,8	66,8	65,2	82,8
	151	90,6	90,4	90,6	88,8	94,8	95,0	93,6	98,2	64,2	64,6	64,0	83,6
50	26	99,6	99,6	99,6	99,4	99,4	99,8	99,2	100	85,4	85,8	85,2	95,0
	81	99,8	99,8	99,8	99,8	100	100	100	100	88,2	88,2	87,8	96,0
	151	99,6	99,8	99,6	99,8	100	100	100	100	88,8	89,6	88,8	96,2

- Kanadyjskie dane pogodowe to obserwacje dziennych temperatur i opadów z 35 kanadyjskich stacji pogodowych (uśrednione w latach 1960-1994, 365 dni). Są one dostępne w pakiecie `fda` programu R.
- Piętnaście stacji pogodowych znajduje się we wschodniej Kanadzie, kolejne piętnaście w zachodniej Kanadzie, a pozostałe pięć w północnej Kanadzie. Poniższy rysunek przedstawia krzywe temperatury i opadów dla tych 35 stacji pogodowych.
- Wydaje się, że temperatury i opady notowane w stacjach północnych są zwykle niższe niż tej w stacjach wschodnich i zachodnich. Przyczyną tego jest zapewne położenie północnych stacji na wyższych szerokościach geograficznych.
- Zatem lokalizacja ma wpływ na średnią temperaturę i opady notowane we wschodnich, zachodnich i północnych stacjach pogodowych. Chcielibyśmy to potwierdzić statystycznie. Dane pogodowe ze stacji wschodnich, zachodnich i północnych to trzy grupy dwuwymiarowych danych funkcjonalnych ($n_1 = n_2 = 15, n_3 = 5, l = 3$).

Przykłady ilustracyjne na danych rzeczywistych



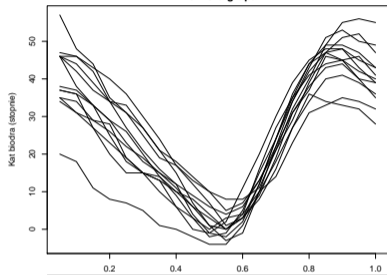
- P-wartości testów permutacyjnych opartych o reprezentację bazową oraz testów opartych o losowe projekcje ($k = 1, 5, 15, 30$) dla kanadyjskich danych pogodowych.

Permutacyjne		Projekcyjne								
W	0	Wp	1	8,82E-10	5	4,27E-07	15	9,19E-07	30	2,54E-08
LH	0	LHp	1	7,89E-11	5	8,46E-08	15	1,35E-08	30	1,57E-08
P	0	Pp	1	1,21E-08	5	3,37E-06	15	1,49E-05	30	4,45E-08
R	0	Rp	1	2,53E-09	5	3,72E-07	15	4,22E-08	30	6,01E-07

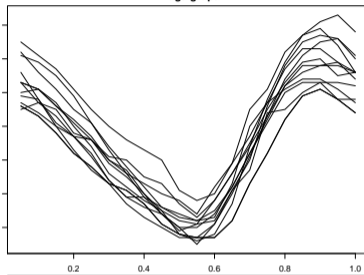
- Dane dotyczące chodu (ang. gait data) składają się z kątów utworzonych przez biodro i kolano każdego z 39 dzieci. Ten zestaw danych znajduje się w pakiecie `fda`. Obserwowane funkcje są równoczesną zmianą kąta biodra i kolana dla 39 dzieci w 20 równomiernie rozłożonych punktach czasowych w $[0,025; 0,975]$. Oba zestawy funkcji są okresowe.
- W celach ilustracyjnych tworzymy trzy grupy tak, że pierwsza grupa składa się z obserwowanych krzywych dla pierwszych 13 dzieci, drugiej grupy krzywych dla kolejnych 13 dzieci i trzeciej grupy krzywych dla pozostałych dzieci. Obserwowane krzywe tych trzech grup przedstawiono na poniższym rysunku. Krzywe obu zmiennych funkcjonalnych wydają się być podobne we wszystkich grupach. Dlatego bardzo możliwe, że średnie krzywe chodu nie różnią się dla trzech grup dzieci. Sprawdzamy to statystycznie, korzystając z proponowanych testów dla trzech prób danych funkcjonalnych, każda o rozmiarze 13.

Przykłady ilustracyjne na danych rzeczywistych

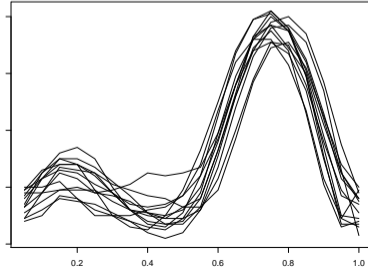
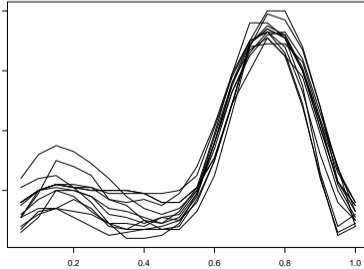
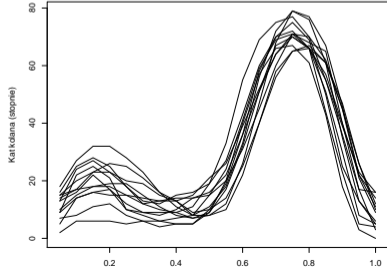
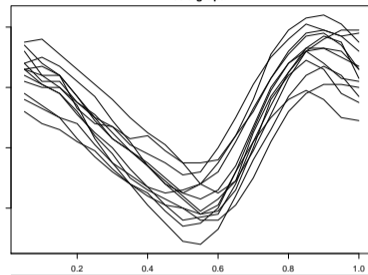
Pierwsza grupa



Druga grupa



Trzecia grupa



- P-wartości testów permutacyjnych opartych o reprezentację bazową oraz testów opartych o losowe projekcje ($k = 1, 5, 15, 20$) dla danych dotyczących chodu.

Permutacyjne		Projekcyjne								
W	0,511	Wp	1	0,256	5	0,295	15	0,276	20	0,281
LH	0,504	LHp	1	0,257	5	0,295	15	0,277	20	0,281
P	0,519	Pp	1	0,257	5	0,295	15	0,276	20	0,281
R	0,379	Rp	1	0,080	5	0,093	15	0,089	20	0,089

- Pakiet `fdANOVA` dostępny w repozytorium CRAN zawiera funkcje realizujące testy analizy wariancji dla jednowymiarowych i wielowymiarowych danych funkcjonalnych.
- `fanova.tests()` - 12 testów FANOVA
- `fmanova.ptbfr()` - testy permutacyjne oparte o reprezentację bazową (różne bazy, kryteria i metody wyboru długości rozwinięcia w bazie)
- `fmanova.trp()` - testy oparte o losowe projekcje (różne liczby losowych projekcji, procesów gaussowskich potrzebnych do projekcji, standardowe i permutacyjne testy na projekcjach)
- `plotFANOVA()` - wykresy jednowymiarowych danych funkcjonalnych z i bez podziału na grupy
- Najbardziej czasochłonne fragmenty programów (np. procedury permutacyjne) zostały zrównoleglone za pomocą pakietu `doParallel`.

- Pakiet [mfds](#) dostępny na GitHub zawiera piętnaście zbiorów etykietowanych i wielowymiarowych danych funkcjonalnych.
- Zbiory te powstały na podstawie wielowymiarowych szeregów czasowych dostępnych w literaturze o różnej liczbie zmiennych, punktów czasowych, grup oraz obserwacji.
- Pochodzą one z różnych dziedzin, w tym rozpoznawanie pisma odręcznego, medycyny, robotyki, itp.
- Zestawy danych można wykorzystać do zilustrowania i praktycznej oceny skuteczności metod klasyfikacji i wnioskowania statystycznego.

- 1 Anderson T.W. (2003) An introduction to multivariate statistical analysis, 3rd edition. Wiley, London.
- 2 Benjamini Y., Yekutieli D. (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 29, 1165–1188.
- 3 Cuesta-Albertos J.A., Febrero-Bande M. (2010) A simple multiway ANOVA for functional data. *Test* 19, 537–557.
- 4 Górecki T., Smaga Ł. (2015) A comparison of tests for the one-way ANOVA problem for functional data. *Computational Statistics* 30, 987–1010.
- 5 Górecki T., Smaga Ł. (2017a) Multivariate analysis of variance for functional data. *Journal of Applied Statistics* 44, 2172–2189.
- 6 Górecki T., Smaga Ł. (2017b) fdANOVA: Analysis of variance for univariate and multivariate functional data. R package version 0.1.0, <http://CRAN.R-project.org/package=fdANOVA>
- 7 Górecki T., Smaga Ł. (2017c) mfdsets: Multivariate functional data sets. R package version 0.1.0, <https://github.com/Halmaris/mfdsets>