

# Vzdělávání a demokratické instituce: analýza kanonických korelací

## 1 Úvod a výzkumné otázky

---

Jednou z opakovaně diskutovaných hypotéz v komparativní politické vědě je teze, že vzdělání a demokratické instituce jsou vzájemně propojené. Vzdělanější populace bývá více politicky angažovaná, lépe odolává dezinformacím a autoritářským tendencím a efektivněji monitoruje veřejný sektor. Méně prozkoumána zůstává otázka, zda tato logika platí i na agregátní úrovni zemí: zda státy s kvalitnějším vzdělávacím systémem a vyšším vzdělanostním kapitálem vykazují zároveň silnější demokratické instituce a nižší korupci.

Tato zpráva zkoumá zmíněný vztah prostřednictvím analýzy kanonických korelací (CCA). Metoda umožňuje analyzovat vztah dvou konceptuálně odlišných sad proměnných jako celků — na rozdíl od sériové regrese, která by vyžadovala tolik modelů, kolik je závislých proměnných, a vedla by ke kumulaci chyby I. druhu a k opomenutí sdílené multivariátní struktury.

*Výzkumná otázka:* Jaký je vztah mezi profilem vzdělávání zemí (výsledky PISA, výdaje na vzdělání, délka vzdělávání) a jejich demokratickým a institucionálním profilem (korupce, demokracie, svoboda), a jakými kanonickými funkcemi lze tento vztah nejlépe popsat?

**H1:** Existuje statisticky významná kanonická korelace mezi sadou vzdělávání (A) a sadou demokratických institucí (B), testovaná Bartlettovým  $\chi^2$  testem ( $\alpha = 0,05$ ).

**H2:** První kanonická funkce bude dominantně sycena výsledky PISA a délkou vzdělávání (sada A) a indexem vnímání korupce a demokratickými ukazateli (sada B).

## 2 Data a metoda

---

### 2.1 Datový soubor a proměnné

Analýza pracuje s průřezovým souborem 67 suverénních států se kompletními daty ze všech zdrojů za rok 2022 (resp. 2021 u výdajů na vzdělání). Subregionální ekonomiky zapojené do PISA 2022 — Hongkong, Macao, Tchaj-wan, Kosovo a Palestinská autonomie — byly z analýzy vyřazeny, aby byl zachován charakter srovnání suverénních zemí. Vzorek zahrnuje státy ze všech světových oblastí a pokrývá celé pásmo výsledků PISA i celou škálu demokratické konsolidace, od plných demokracií po autoritářské režimy.

**Sada A — vzdělávání** zachycuje tři dimenze. Průměrné skóre PISA 2022 (PISA\_avg) je aritmetický průměr výsledků 15letých žáků v matematice, čtenářské gramotnosti a přírodních vědách, jak jej publikuje OECD. Výdaje na vzdělání (Educ\_exp) představují vládní výdaje na vzdělání jako procento HDP podle dat Světové banky. Průměrná délka vzdělávání (Schooling) zachycuje průměrný počet let vzdělávání populace starší 25 let podle Human Development Report UNDP 2022.

**Sada B — demokratické instituce** zahrnuje rovněž tři proměnné. Index vnímání korupce (CPI 2022) publikovaný organizací Transparency International nabývá hodnot 0 (maximální korupce) až 100 (minimum korupce). Index demokracie EIU (DemIdx 2022) hodnotí každou zemi na škále 0–10 na základě pěti komponent: volební procesy, fungování vlády, politická participace, politická kultura a občanské svobody. Souhrnné skóre Freedom House (FH 2022) zachycuje míru politické svobody a občanských práv na škále 0–100. Přehled proměnných je v Tabulce 1.

Tabulka 1 Přehled proměnných, zdrojů a škál.

Proměnná	Zdroj / rok	Rozsah	Operacionalizace	Sada
PISA_avg	OECD PISA 2022	350–560	Průměr skóre mat. / čtenář. / věda	A
Educ_exp	World Bank, 2021	2,6–7,8 %	Výdaje na vzdělání, % HDP	A
Schooling	UNDP HDR 2022	5,9–14,1 let	Průměrná délka vzdělávání 25+	A
CPI	TI CPI 2022	0–100	Index vnímání korupce (↑ = méně korupce)	B
DemIdx	EIU 2022	0–10	Index demokracie	B
FH	Freedom House 2022	0–100	Svoboda voleb, tisku, práv (↑ = svobodnější)	B

## 2.2 Statistická metoda

Analýza kanonických korelací (CCA) je vícerozměrná technika, která hledá lineární kombinace proměnných v sadě A — tzv. kanonický variát  $u$  — a lineární kombinace proměnných v sadě B — kanonický variát  $v$  — tak, aby jejich vzájemná (kanonická) korelace  $rc$  byla maximální. Výsledkem je  $k = \min(p, q) = \min(3, 3) = 3$  kanonické funkce, přičemž každá zachycuje dimenzi vztahu obou sad nezávislou na ostatních funkcích. Při interpretaci se pozornost soustředí na statisticky průkazné kanonické páry.

Vhodnost CCA oproti sériové regresi nebo bivariantní korelační analýze spočívá v tom, že metoda simultánně optimalizuje vztah obou sad jako celků. To zabraňuje kumulaci chyby I. druhu, která by vznikla opakovaným testováním, a zároveň zachycuje sdílené multivariátní vzorce, které by bivariantní přístupy přehlédly. Klíčovým předpokladem je přítomnost věcně smysluplných korelací uvnitř každé ze sad i mezi nimi — viz Oddíl 3.1 a Obrázek 1.

Statistická významnost kanonických korelací byla testována sekvenčním Bartlettovým  $\chi^2$  testem: nejprve pro všechny tři funkce simultánně, poté pro zbývající dvě po vyřazení první a nakonec pro třetí samotnou. Wilksovo lambda ( $\Lambda$ ) zachycuje podíl nevysvětlené variance. Míra věcné velikosti vztahu je vyjádřena indexem redundance (Stewart & Love, 1968), který udává, jaké procento variance jedné sady proměnných je vysvětleno druhou sadou prostřednictvím daného kanonického páru. Výpočty byly provedeny v jazyce Python 3 s využitím knihoven NumPy a SciPy; kanonická dekompozice byla realizována metodou singulárního rozkladu (SVD) po standardizaci obou sad.

Před provedením CCA byla ověřena splnění klíčových předpokladů. Linearita vztahu mezi sadami byla posouzena graficky z rozptylového diagramu prvního kanonického páru. Multikolinearita uvnitř sad byla ověřena podmínkovým číslem kovariančních matic:  $S_{xx} = 5,3$  a  $S_{yy} = 8,1$ , obě výrazně pod kritickým prahem 30 (Tabachnick & Fidell, 2019). Mírné odchylky od normality (zejm. bimodální distribuce CPI a FH) jsou při  $n = 67$  akceptovatelné — CCA je vůči nim relativně robustní.

## 3 Výsledky

### 3.1 Deskriptivní statistika a korelace

Tabulka 2 shrnuje deskriptivní charakteristiky všech šesti proměnných. Průměrné PISA skóre ( $M = 447,3$ ;  $SD = 49,4$ ) je mírně pod průměrem zemí OECD, což odráží zastoupení rozsáhlé skupiny rozvíjejících se zemí. Výdaje na vzdělání vykazují relativně malý rozptyl ( $M = 4,7$  %;  $SD = 1,2$  %) a — jak ukazuje Tabulka 3 — korelují s výsledky PISA jen slabě ( $r = 0,23$ ), naznačující, že výše výdajů sama o sobě vzdělávací výsledky nezajišťuje.

Tabulka 2 Deskriptivní statistika (n = 67).

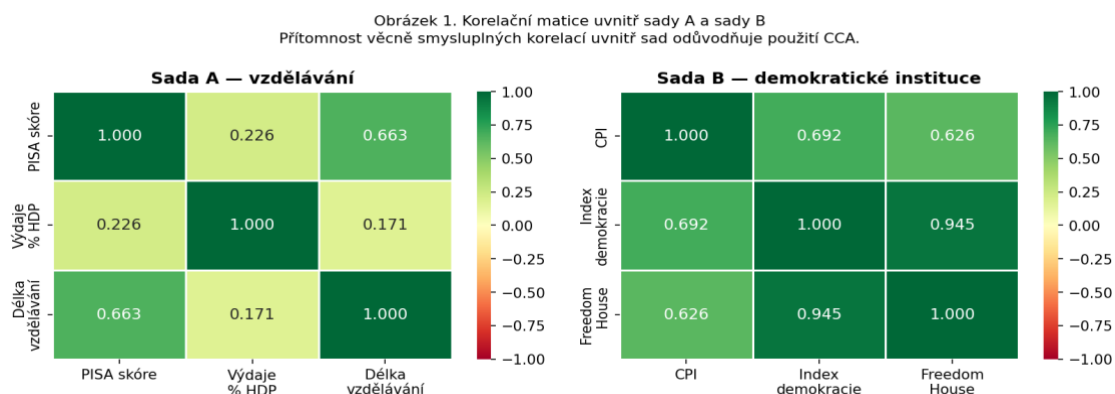
	PISA_avg	Educ_exp	Schooling	CPI	Demldx	FH
M	447,3	4,72	11,27	56,2	6,91	73,9
SD	49,4	1,18	1,77	17,6	1,89	25,7
Min	352,7	2,6	5,9	23	2,08	7
Mdn	459,0	4,70	11,70	55	7,34	83
Max	559,7	7,8	14,1	90	9,81	100

Tabulka 3 zobrazuje korelační matici všech šesti proměnných. V sadě A jsou vnitřní korelace mírné: PISA\_avg a Schooling ( $r = 0,663$ ), PISA\_avg a Educ\_exp ( $r = 0,226$ ), Schooling a Educ\_exp ( $r = 0,171$ ). V sadě B jsou korelace výrazně vyšší — Demldx a FH sdílejí téměř totožný rozptyl ( $r = 0,945$ ), CPI a Demldx ( $r = 0,692$ ). Mezišetové korelace (označeny ★) jsou střední až silné ( $r = 0,23-0,74$ ), čímž je naplněn základní předpoklad CCA: proměnné obou sad musejí spolu věcně smysluplně korelovat, aby CCA dávala smysl.

Tabulka 3 Korelační matice všech šesti proměnných. ★ = mezišetové korelace (A × B). Horní trojúhelník.

	PISA_avg	Educ_exp	Schooling	CPI	Demldx	FH
PISA_avg	—	0,226	0,663	0,740★	0,573★	0,536★
Educ_exp		—	0,171	0,426★	0,440★	0,415★
Schooling			—	0,628★	0,529★	0,552★
CPI				—	0,692	0,626
Demldx					—	0,945
FH						—

Korelační matice uvnitř jednotlivých sad jsou vizualizovány na Obrázku 1. Srovnání obou panelů ukazuje klíčový strukturální rozdíl: zatímco korelace uvnitř sady A jsou nízké až střední (proměnné zachycují relativně odlišné aspekty vzdělávání), korelace uvnitř sady B jsou silné až velmi silné (Demldx a FH v podstatě měří tentýž konstrukt ze dvou úhlů). Tato skutečnost je brána v úvahu při interpretaci strukturálních koeficientů v Oddíle 3.3.



Obrázek 1. Korelační matice proměnných uvnitř sady A (vzdělávání, vlevo) a sady B (instituce, vpravo). Silné mezišetové korelace ( $r = 0,23-0,74$ ) odůvodňují použití CCA; vysoká vnitřní korelace Demldx–FH ( $r = 0,945$ ) bude brána v úvahu při interpretaci.

### 3.2 Bartlettovy testy — výsledky CCA

Výsledky sekvenčního Bartlettova testu jsou v Tabulce 4. Test pro všechny tři kanonické funkce simultánně dosáhl vysoké statistické průkaznosti: Wilks  $\Lambda = 0,293$ ,  $\chi^2(9) = 76,77$ ,  $p < 0,001$ . H1 je potvrzena — mezi sadami A a B existuje statisticky velmi významná kanonická korelace. Po vyřazení první kanonické funkce zbývající dvě nejsou statisticky průkazné ( $p = 0,34$  resp.  $p = 0,27$ ). Celý věcně smysluplný vztah obou sad je tedy zachycen jedinou kanonickou dimenzí — vztah je jednorozměrný, nikoli vícevrstevnatý.

Tabulka 4 Sekvenční Bartlettovy  $\chi^2$  testy. \*\*\*  $p < 0,001$ . rc = kanonická korelace,  $rc^2$  = sdílená variance páru,  $\Lambda$  = Wilksovo lambda pro zbývající funkce.

Funkce	rc	$rc^2$	Wilks $\Lambda$	$\chi^2$	df	p	Závěr
1–3	0,828	0,685	0,293	76,77	9	< 0,001	*** sig.
2–3	0,228	0,052	0,930	4,53	4	0,339	ns
3	0,139	0,019	0,981	1,21	1	0,271	ns

### 3.3 Strukturní koeficienty a interpretace CV1

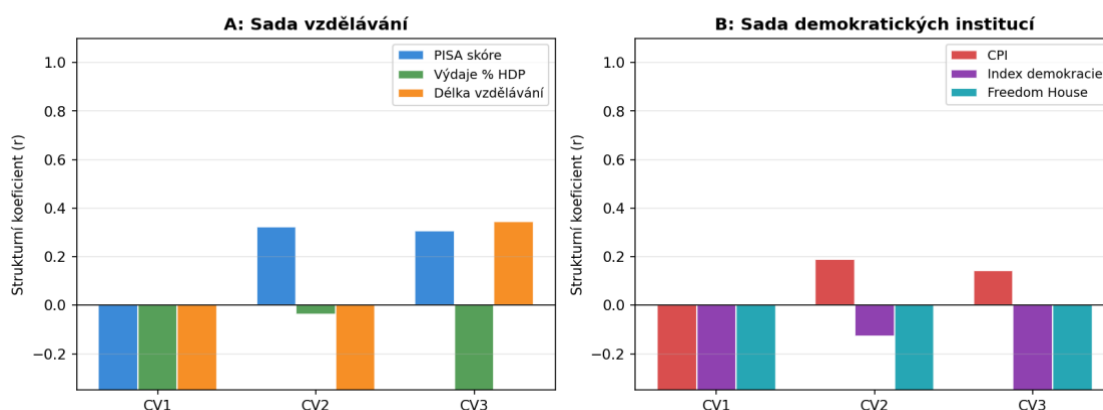
Strukturní koeficienty (kanonické zátěže) vyjadřují korelaci každé původní proměnné s příslušným kanonickým variátem a jsou klíčovým nástrojem pro věcnou interpretaci výsledků CCA. Jsou zobrazeny v Tabulce 5 a na Obrázku 2. Pozornost se soustředí výhradně na první kanonickou funkci (CV1), neboť pouze ta je statisticky průkazná.

V sadě A jsou strukturní koeficienty pro CV1 záporné: PISA\_avg ( $r_s = -0,895$ ), Schooling ( $r_s = -0,799$ ), Educ\_exp ( $r_s = -0,552$ ). Záporné znaménko je konvenční — odvíjí se od matematické orientace kanonického variátu, nikoli od věcného směru vztahu. Věcně lze říci, že CV1 v sadě A je dominantně tvořen výsledky PISA a délkou vzdělávání, přičemž výdaje na vzdělání přispívají o poznání méně. V sadě B jsou zátěže CV1 rovněž záporné a silné: CPI ( $r_s = -0,972$ ), DemIdx ( $r_s = -0,819$ ), FH ( $r_s = -0,792$ ). Všechny tři institucionální proměnné tvoří kohezní kanonický variát, přičemž index vnímání korupce je jeho nejsilnější složkou. H2 je potvrzena.

Tabulka 5 Strukturní koeficienty (kanonické zátěže) pro tři kanonické funkce. Interpretovány jsou zátěže CV1 jako jediné statisticky průkazné funkce.

Proměnná	Sada	$r_s$ (CV1)	$r_s$ (CV2)	$r_s$ (CV3)
PISA_avg	A	-0,895	+0,323	+0,307
Schooling	A	-0,799	-0,492	+0,346
Educ_exp	A	-0,552	-0,037	-0,833
CPI	B	-0,972	+0,190	+0,142
DemIdx	B	-0,819	-0,129	-0,559
FH	B	-0,792	-0,434	-0,429

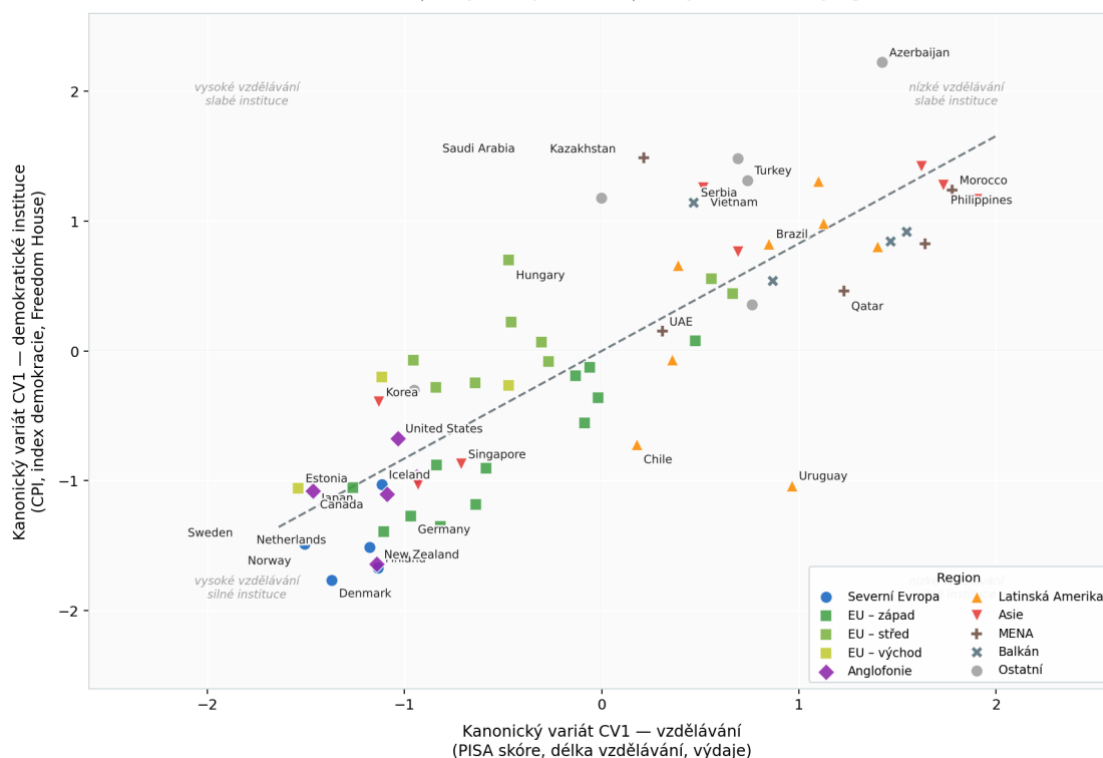
Obrázek 2. Strukturální koeficienty (kanonické zátěže) pro sady A a B  
 $rc_1 = ,828^{***}$ ,  $rc_2 = ,228$  ns,  $rc_3 = ,139$  ns — interpretovatelná je pouze CV1



Obrázek 2. Strukturální koeficienty proměnných v sadě A (vzdělávání, vlevo) a sadě B (instituce, vpravo) pro tři kanonické funkce. Záporné hodnoty CV1 jsou konvenční — věcně jsou všechny proměnné s CV1 pozitivně asociovány.

CV1 zachycuje dimenzi, kterou lze pojmenovat jako "vzdělávací a institucionální kapitál země". Na záporném pólu se shlukují země s vysokými výsledky PISA, vzdělanou populací, nízkou korupcí a silnou demokracií — typicky severské státy, anglofonní demokracie a Estonsko. Na kladném pólu se nacházejí země s nižšími výsledky PISA, kratší průměrnou dobou vzdělávání a vyšší korupcí — Azerbajdžán, Kazachstán, Saúdská Arábie nebo Maroko. Tato interpretace je ilustrována rozptylovým diagramem na Obrázku 3.

Obrázek 3. První kanonický pár (CV1 × CV1),  $rc_1 = ,828$ ,  $p < ,001$ ,  $rc_1^2 = ,685$   
 $n = 67$  zemí | každý bod = jedna země | tvar symbolu = světový region



Obrázek 3. Rozptylový diagram prvního kanonického páru CV1 × CV1 ( $rc_1 = 0,828$ ,  $p < 0,001$ ,  $rc_1^2 = 0,685$ ). Každý bod = jedna země; tvar symbolu = světový region. Přerušovaná čára = regresní přímka. Popisky kvadrantů znázorňují věcnou interpretaci os.

### 3.4 Index redundance

Data a další informace o této zprávě jsou dostupné na adrese <https://dostal.vyzkum-psychoologie.cz/stat4?i=840>

Index redundance (Stewart & Love, 1968) udává, jakou část celkové variance v jedné sadě vysvětlují kanonické variáty druhé sady. Prostřednictvím CV1 je vysvětleno 51,2 % sdílené variance sady B (instituce) sadou A (vzdělávání). Celkový redundanční index pro sadu B činí 52,0 % — zbývající dvě kanonické funkce přidávají dohromady méně než 1 %. Pro sadu A je redundanční index 39,9 % prostřednictvím CV1 (celkově 41,0 %). Asymetrie v hodnotách (52 % vs. 41 %) naznačuje, že vzdělávací profil zemí predikuje jejich institucionální kvalitu o něco silněji než naopak — avšak symetrie CCA brání jakýmkoli kauzálním interpretacím.

## 4 Diskuze

---

Klíčovým nálezem zprávy je silná, statisticky vysoce průkazná první kanonická korelace ( $rc_1 = 0,828$ ,  $p < 0,001$ ,  $rc_1^2 = 0,685$ ) a jednorozměrnost vztahu obou sad. Vzdělávací a demokratický profil zemí jsou tedy v podstatě dvěma facetami téhož jevu — na úrovni zemí jde o vzdělávání a demokratické instituce systematicky ruku v ruce. Výsledek je v souladu s empirickou literaturou: Glaeser et al. (2004) demonstrují, že vzdělání je silnějším prediktorem demokratické konsolidace než HDP per capita, a Acemoglu et al. (2005) oba faktory interpretují jako vzájemně se posilující v rámci institucionálních podmínek.

Pozoruhodná je dominující role CPI v sadě B ( $r_s = -0,972$ ). Index vnímání korupce strukturuje institucionální prostor silněji než čistě politické ukazatele jako DemIdx či FH. Nízká korupce je zároveň podmínkou fungujícího vzdělávacího systému — systém bez korupce alokuje zdroje efektivněji, zajišťuje meritokratický přístup a zabraňuje klientelismu — i jeho potenciálním výsledkem, neboť vzdělaná populace lépe monitoruje veřejný sektor a vytváří poptávku po transparentnosti institucí.

Interpretačně zajímavá je slabší zátěž Educ\_exp ( $r_s = -0,552$ ) ve srovnání s PISA\_avg ( $-0,895$ ) a Schooling ( $-0,799$ ). Výdaje na vzdělání jako procento HDP zachycují *vstup* do vzdělávacího systému, nikoli jeho *výstup*. Právě tento rozdíl se v datech zřetelně projevuje: Saúdská Arábie vynakládá 7,8 % HDP — nejvíce ze vzorku — avšak dosahuje PISA skóre pouhých 387 bodů a je jednou z nejautoritářšších zemí (FH = 7). To naznačuje, že pro demokratizaci a institucionální kvalitu jsou relevantní vzdělávací výstupy — tedy to, co studenti skutečně umějí a jak dlouho se vzdělávají — spíše než samotná výše financování.

Zvláštní pozornost zaslouží dva typy zemí dokumentující meze jednorozměrného vztahu. Singapur a Vietnam leží v levé horní části Obrázku 3 — kombinují nadprůměrné vzdělávací výsledky s velmi nízkou demokratickou svobodou (Singapur: PISA = 559,7, FH = 48; Vietnam: PISA = 467,7, FH = 19). Tyto případy demonstrují, že autoritářský režim může dosáhnout vysokých vzdělávacích výsledků prostřednictvím silné státní intervence do školství bez nutnosti demokratické konsolidace. Opačný pól ilustruje Uruguay (PISA = 404,3, FH = 97, DemIdx = 8,30): kombinuje spíše průměrné vzdělávací výsledky s plnou demokratickou svobodou, čímž ukazuje, že demokratická kultura může být hluboce zakořeněna i bez výjimečných vzdělávacích výsledků.

Jednorozměrnost výsledků — druhá a třetí kanonická funkce jsou statisticky nevýznamné ( $rc_2 = 0,228$ ,  $p = 0,34$ ;  $rc_3 = 0,139$ ,  $p = 0,27$ ) — je interpretačně čistým výsledkem. V datech neexistují další zachytitelné dimenze vztahu vzdělávání a demokratických institucí. To neznámá, že takové dimenze neexistují: s větším vzorkem zahrnujícím i státy bez PISA dat by druhý kanonický pár mohl zachytit skupinu zemí s vysokými výdaji na vzdělání, ale nízkou demokratickou kvalitou.

## 5 Limity a závěr

---

Tato studie prokázala statisticky vysoce průkaznou a věcně silnou kanonickou korelaci ( $rc_1 = 0,828$ ,  $p < 0,001$ ) mezi profilem vzdělávání zemí a jejich demokraticko-institucionálním profilem. Vztah je jednorozměrný a zachycuje osu, na níž stojí na jednom konci severské státy a anglofonní demokracie, na druhém autoritářské státy s nižšími vzdělávacími výsledky.

Data a další informace o této zprávě jsou dostupné na adrese <https://dostal.vyzkum-psychologie.cz/stat4?i=840>

Redundanční analýza ukazuje, že vzdělávací profil zemí vysvětluje 52 % sdílené variance v institucionálním profilu.

Výsledky je třeba interpretovat s ohledem na několik limitů. Za prvé, průřezový design analýzy nedovoluje vyvozovat kauzální závěry. Za druhé, výsledky platí na úrovni zemí; přímé zobecnění na individuální úroveň by bylo zatíženo ekologickou falací. Za třetí, výběr vzorku je podmíněn účastí v PISA 2022, která systematicky vylučuje nejchudší státy světa. Za čtvrté, vysoká korelace DemIdx a FH ( $r = 0,945$ ) naznačuje překrývání měřených konstruktů; jejich nahrazení méně korelovanými proměnnými (např. indexem právního státu World Bank) by mohlo přinést diferencovanější pohled. Přes tato omezení přináší analýza jasný a robustní závěr: vzdělávání a absence korupce jdou na úrovni zemí systematicky ruku v ruce, a to v celém rozsahu globální diverzity zkoumaného vzorku.

## 6 Reference

---

- Acemoglu, D., Johnson, S., Robinson, J. A., & Yared, P. (2005). From education to democracy? *American Economic Review*, 95(2), 44–49. <https://doi.org/10.1257/000282805774669916>
- Economist Intelligence Unit. (2023). *Democracy Index 2022: Frontline democracy and the battle for Ukraine*. <https://www.eiu.com/n/campaigns/democracy-index-2022/>
- Freedom House. (2022). *Freedom in the World 2022: The global expansion of authoritarian rule*. <https://freedomhouse.org/report/freedom-world/2022/global-expansion-authoritarian-rule>
- Glaeser, E. L., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2004). Do institutions cause growth? *Journal of Economic Growth*, 9(3), 271–303. <https://doi.org/10.1023/B:JOEG.0000038933.16398.ed>
- OECD. (2023). *PISA 2022 Results (Volume I): The state of learning and equity in education*. <https://doi.org/10.1787/53f23881-en>
- Stewart, D. K., & Love, W. A. (1968). A general canonical correlation index. *Psychological Bulletin*, 70(3), 160–163. <https://doi.org/10.1037/h0026143>
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.
- Transparency International. (2023). *Corruption Perceptions Index 2022*. <https://www.transparency.org/en/cpi/2022>
- UNDP. (2022). *Human Development Report 2021/2022: Uncertain times, unsettled lives*. <https://hdr.undp.org/content/human-development-report-2022>
- World Bank. (2023). *Government expenditure on education, total (% of GDP)* [Data set]. <https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS>