

# Vícenásobná lineární regrese akademického výkonu studentů

## Úvod

Výsledky studentů jsou dány souhrou mnoha faktorů, mezi něž patří jak studijní návyky, tak životní styl. Výzkumy ukazují, že kvalita a efektivita času stráveného studiem, stejně jako dostatečný spánek, jsou klíčové pro udržení optimální kognitivní funkce a dosažení dobrých výsledků. Například Plant et al. (2005) argumentují, že nejen samotná doba strávená učením, ale především způsob, jakým studenti studují, významně ovlivňuje jejich akademický úspěch. Podobně Curcio et al. (2006) poukazují na negativní dopady nedostatku spánku, který snižuje schopnost učení a celkový výkon. Také Trockel et al. (2000) zjistili, že spánkové vzorce a další zdravotně zaměřené návyky studentů významně korelují s jejich studijními výsledky.

Na začátek je důležité zmínit, že data, z nichž tato analýza vychází byla vytvořena uměle a nejedná se o výsledky skutečné studie. Zdrojem dat byla vědecká knihovna kaggle.com. Cílem této fiktivní studie bylo zjistit, do jaké míry lze předpovědět výkonnost studentů pomocí tří faktorů: předchozích výsledků, počtu hodin strávených studiem a průměrné doby spánku. Výkonnost studentů, vyjádřená koeficientem  $V$ , se pohybovala v intervalu od 10 do 100. Pro modelování vztahu mezi proměnnými byla využita vícenásobná lineární regrese.

## Metoda

Regresní model byl specifikován následující rovnicí:

$$V = \beta_A \cdot A + \beta_B \cdot B + \beta_C \cdot C + \beta_0,$$

- $\beta_A$  až  $\beta_C$  jsou regresní koeficienty udávající změnu výkonnosti  $V$  při jednotkové změně odpovídající proměnné za předpokladu, že ostatní proměnné zůstanou beze změny,
- $A$  reprezentuje předchozí výsledky studentů,
- $B$  je počet hodin věnovaných studiu,
- $C$  udává průměrnou dobu spánku,
- $\beta_0$  je hodnota počátku neboli interceptu.

Model byl odhadnut na základě 10 000 záznamů. Pro posouzení kvality modelu byl vyhodnocen koeficient determinance ( $R^2$ ), koeficient mnohonásobné korelace ( $R$ ), F-test a statistická významnost jednotlivých regresorů pomocí Waldovy statistiky.

## Výsledky

Model dosáhl velmi vysoké kvality, což potvrzují následující statistické ukazatele. Koeficient determinance  $R^2$  dosáhl hodnoty 0,9876, což znamená, že model vysvětluje přibližně 98,76 % variability výkonnosti studentů. Velmi vysoká hodnota koeficientu mnohonásobné korelace  $R = 0,9938$  poukazuje na silnou souvislost mezi předpovědí modelu a skutečnými hodnotami závislé proměnné.

Data a další informace o této zprávě jsou dostupné na adrese <https://dostal.vyzkum-psychologie.cz/stat4?i=502>.

Testová statistika  $F(3,996) = 266461$  s p-hodnotou  $< 0,0000$  potvrzuje, že alespoň jeden z regresorů významně přispívá k vysvětlení variability výkonového indexu. Nízká směrodatná odchylka predikce (2,135) ukazuje na vysokou přesnost modelu. Všechny tři proměnné se ukázaly být jako statisticky významné. Níže uvedená tabulka shrnuje výsledky modelu.

Proměnná	Nestandardizovaný koeficient $\beta$	Standardizovaný koeficient $\beta^*$	Waldova statistika	p-hodnota
A	1,019	0,920	827,3	0,00
B	2,857	0,385	346,4	0,00
C	0,478	0,042	37,9	0,00

V tomto modelu vychází počátek na hodnotu -32,915. Jeho interpretace tak může být na první pohled překvapivá. Je však důležité si uvědomit, že počátek udává teoretickou hodnotu závislé proměnné v případě, že všechny prediktory nabývají nulových hodnot. V kontextu této studie by to znamenalo situaci, kdy student nemá žádné předchozí výsledky, netráví žádný čas studiem a vůbec nespí, což je samozřejmě nereálný scénář. Počátek je nutné vnímat jako hodnotu, která kalibruje model, a pomáhá posunout predikce do reálného rozsahu hodnot. Výsledná rovnice predikující výkonnost studentů má tedy následující tvar:

$$V = 1,019 \cdot A + 2,857 \cdot B + 0,478 \cdot C - 32,915.$$

## Diskuse a závěr

Vícenásobná lineární regrese naznačuje, že kombinace předchozích výsledků, počtu hodin studia a průměrné doby spánku velmi přesně předpovídá výkonnost studentů. Vysoká hodnota  $R^2$  (98,76 %) ukazuje, že téměř veškerá variabilita je vysvětlena těmito třemi proměnnými. Hodnota testové statistiky F a p-hodnota potvrzují, že model je statisticky významný. Směrodatná odchylka predikce je také poměrně nízká, což svědčí o přesnosti modelu.

Jednoznačně největší vliv mají předchozí výsledky jejichž standardizovaný koeficient měl hodnotu 0,920, což z této proměnné dělá klíčový prediktor. Druhým nejsilnějším prediktorem se ukázala být doba studia. Její standardizovaný koeficient dosáhl hodnoty 0,385. Z výsledků plyne, že každá hodina studia zvyšuje výkonost o 2,857 bodů. Spánek má sice statisticky významný, ale relativně zanedbatelný vliv (standardizovaný koeficient 0,042). Hodina spánku přináší nárůst o 0,478 bodu.

Model vícenásobné lineární regrese tedy ukázal, že předchozí skóre a počet hodin strávených studiem jsou silnými prediktory akademického výkonu, zatímco průměrný počet hodin spánku má menší, avšak statisticky významný vliv. Celkově model vysvětluje téměř 99 % variability výkonového indexu, což z něj činí velmi robustní nástroj pro předpověď akademického úspěchu na základě zvolených faktorů.

Data a další informace o této zprávě jsou dostupné na adrese <https://dostal.vyzkum-psychologie.cz/stat4?i=502>.

## Zdroje

Curcio, G., Ferrara, M., & De Gennaro, L. (2006). Sleep loss, learning capacity and academic performance. *Sleep Medicine Reviews*, 10(5), 323–337. <https://doi.org/10.1016/j.smr.2005.11.001>.

Plant, E. A., Ericsson, K. A., Hill, L., & Asberg, K. (2005). Why study time does not predict grade point average across college students: Implications of deliberate practice for academic performance. *Contemporary Educational Psychology*, 30(4), 376–395. <https://doi.org/10.1016/j.cedpsych.2004.06.001>.

Trockel, M. T., Barnes, M. D., & Egget, D. L. (2000). Health-related variables and academic performance among first-year college students: Implications for sleep and other behaviors. *Journal of American College Health*, 49(3), 125–131. <https://doi.org/10.1080/07448480009596294>.