

# Vliv kouření cigaret na výkon výdechu aneb Feather Propelling Contest

## Úvod

Kouření cigaret je již dlouhodobě známé jako faktor negativně ovlivňující zdraví člověka (Sherman, 1991; Jayes et al., 2016). Je prokazatelně spojeno s různými neurologickými, kardiovaskulárními a plicními chorobami (Das, 2003). Z metaanalýzy Yayese a kol. (2014) se ukázalo, že se mezi dospělými kuřáky potvrdili výrazně zvýšená rizika vzniku rakoviny plic, chronická obstrukční plicní nemoc a astmatu.

V rámci této studie se zaměřujeme na analýzu vlivu kuřáckého statusu na výkon ve výdechu prostřednictvím Feather Propelling Contest, což je disciplína, která měří schopnost foukat peříčko, s cílem posunout jej co nejvíce vpřed. Naše zájmové zaměření tkví v identifikaci případného vztahu mezi kouřením cigaret a výsledky této specifické disciplíny.

Zatím nebylo nalezeno mnoho studií, které by se přímo zabývaly tímto tématem. Je však důležité zdůraznit, že některá respirační onemocnění spojená s kouřením cigaret, jako je například chronická obstrukční plicní nemoc, mohou významně ovlivnit fyziologický výkon dýchání. Tato onemocnění zahrnují emfyzém a chronickou bronchitidu, které postupně omezuje průtok vzduchu do plic. Navíc je známo, že astma může také snížit výkon výdechu (Feschenko et al., 2018).

## Princip experimentu, proměnné a zpracování dat

Výzkumná data byla získána z tzv. Feather Propelling Contest, soutěže, během níž účastníci foukali do peříčka a byla měřena jejich maximální dosažená vzdálenost (v metrech), kam peří doputovalo. Soubor dat obsahuje 300 pozorování od 100 účastníků<sup>1</sup>. Každý účastník měl 3 pokusy. Cílem studie bylo prozkoumat, zda existuje vztah mezi kuřáckým status a dosaženým výkonem ve foukání. Zahrnuté proměnné zahrnovaly kuřácký status účastníků (kuřák/nekuřák) a vzdálenost, kterou dosáhli během soutěže. Data byla podrobně analyzována za účelem odhalení případných vzorů nebo rozdílů v závislosti na kuřáckém statusu, především jsme se zaměřili na model se smíšenými efekty a zkoumání jeho validity.

## Lineární model se smíšenými efekty a práce s daty

Pro analýzu dat byl použit lineární model se smíšenými efekty, což je statistická metoda vhodná pro analýzu opakovaných měření nebo dat s hierarchickou strukturou, jako je například soubor dat z naší studie, kde jednotlivé pokusy byly provedeny na úrovni jednotlivých účastníků. Tento model umožňuje zahrnout jak pevné efekty faktorů zájmu (např. kuřácký status), tak i náhodné efekty (např. variabilitu mezi jednotlivými účastníky), což umožňuje lepší kontrolu nad potenciálními zkresleními a získání spolehlivějších odhadů parametrů.

Analýza modelu se zaměřila nejen na samotný vztah mezi kuřáckým statutem a výkonem ve foukání, ale také na zkoumání validity modelu, což je důležitý krok pro zajištění přesnosti a spolehlivosti výsledků. Validita modelu se zabývala otázkami, zda model správně zachycuje reálnou strukturu dat a zda dosažené výsledky lze obecně aplikovat na celkovou populaci účastníků. Tato důkladná analýza přispívá k důvěryhodnosti a interpretovatelnosti výsledků naší studie.

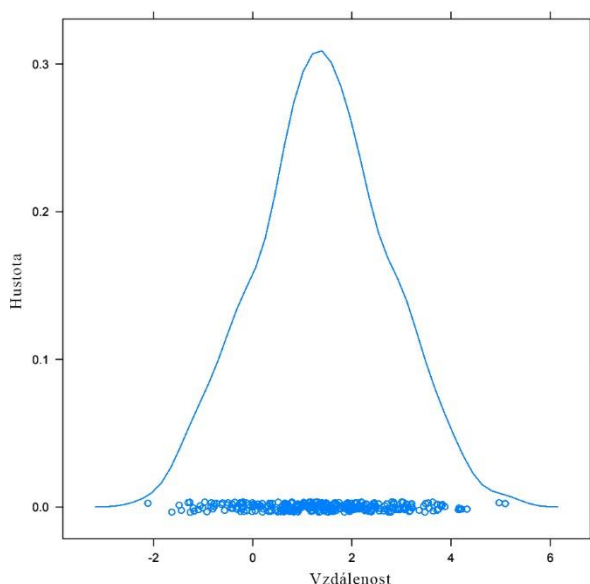
---

<sup>1</sup> Data a další informace o této zprávě jsou dostupné na adrese <https://dostal.vyzkum-psychoologie.cz/stat4?i=268>.

Závislou proměnnou byla dosažená vzdálenost, nezávislou proměnnou kuřácký status účastníka (kuřák/nekuřák). Data jsme převedli na takzvané **vycentrované hodnoty** a dále s nimi pracovali v této podobě. Tyto hodnoty jsou vypočítány jako rozdíl mezi každou původní hodnotou a střední hodnotou celého souboru dat. Tímto způsobem jsme dosáhli toho, že průměr vycentrovaných hodnot je nula. Záporné vycentrované hodnoty pak znamenají, že daná hodnota je nižší než průměr, zatímco kladné hodnoty jsou vyšší než průměr.

Jedním z těchto kroků bylo zkoumání rozložení závislé proměnné, tj. vzdálenosti dosažené účastníky. Vykreslili jsme graf hustoty (Graf 1), který zobrazuje distribuci těchto vzdáleností. Tento graf nám pomohl posoudit, zda data vykazují normální rozdělení, což je jeden z předpokladů pro správné použití lineárního modelu. Hustota v tomto kontextu znamená, jak často se účastníci dostali do různých vzdálenostních kategorií. Z grafu je patrný zvonovitý tvar, který je žádoucí pro určení normálního rozdělení.

**Graf 1:** Graf hustoty zobrazující distribuci dosažených vzdáleností



Také je užitečné zkoumat, jak je užitý model přesný ve svých predikcích. K tomu nám může posloužit například bodový graf zobrazující předpovězené hodnoty modelu oproti skutečným pozorovaným hodnotám, který můžete vidět pod tímto odstavcem.

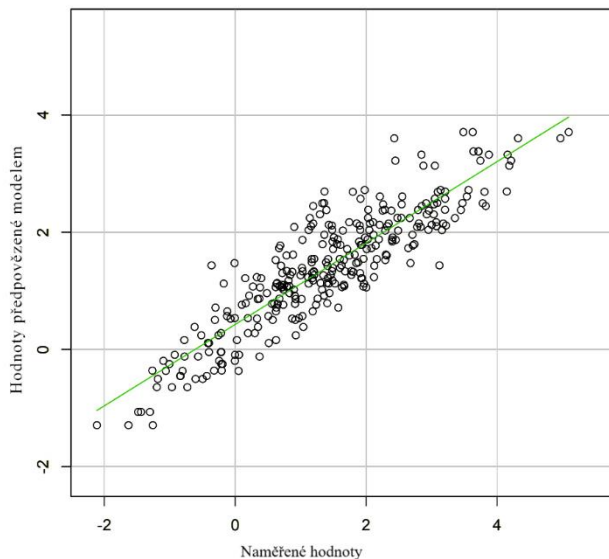
Ideálně bychom chtěli vidět, že body na grafu leží přibližně kolem přímky reprezentující perfektní shodu mezi předpovědí a realitou. To by naznačovalo, že model dokáže dobře vysvětlit rozptyl v datech.

Z grafu vidíme, že pro střední hodnoty vysvětlující proměnné leží body poměrně těsně kolem přímky, což značí dobrou shodu modelu s daty v této oblasti. Nicméně pro extrémně nízké a vysoké hodnoty se předpovědi začínají odchylovat od pozorovaných dat. Zejména pro nejvyšší hodnoty má model tendenci predikce podhodnocovat.

Celkově se tedy zdá, že použitý model relativně dobře popisuje vztah mezi proměnnými pro běžné hodnoty dat. Avšak pro extrémní hodnoty, obzvláště ty nejvyšší, má problémy s přesnou predikcí a mírně je podhodnocuje. Nicméně je možné, že tato tendence modelu

vyhýbat se extrémním predikčním hodnotám může být do jisté míry žádoucí. Umírněnější odhady na koncích rozložení dat mohou zabránit přílišnému ovlivnění modelu odlehlými hodnotami či výjimečnými případy. Takovéto chování může vést k robustnějšímu modelu, který není náchylný k nežádoucím výkyvům způsobeným ojedinělými odchylkami v datech.

**Graf 2:** Bodový graf zobrazující předpovězené hodnoty modelu oproti skutečným pozorovaným hodnotám



Zelená regresní čára je viditelná na grafu. Tato čára představuje lineární regresi mezi dvěma proměnnými: Reálnou vzdáleností (na ose x) a hodnotami předpovězenými modelem (na ose y). Sklon regresní čáry ukazuje, jak se předpovězené hodnoty mění s reálnou vzdáleností, které účastníci dosáhli při soutěži. V našem případě je sklon pozitivní, což naznačuje, že s rostoucí vzdáleností se zvyšují i předpovězené hodnoty.

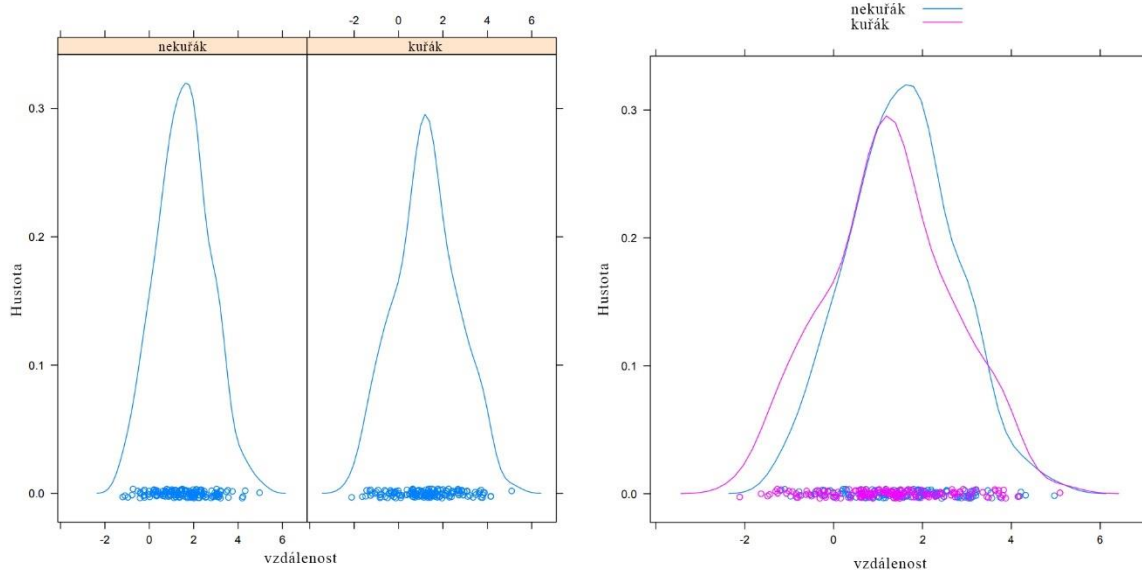
Body na grafu jsou rozptýleny kolem regresní čáry. To znamená, že i když existují odchylky, model se celkově snaží dobře přizpůsobit pozorovaným datům. Některé body jsou blíže regresní čáře, zatímco jiné jsou od ní vzdálenější. To ukazuje, že existuje určitá variabilita v datech.

Analýza rozložení dat a predikce modelu ukázala, že model dobře vystihuje střední hodnoty dosažených vzdáleností. Celkově zvolený přístup poskytuje vhodný rámec pro interpretaci výsledků týkajících se vlivu kuřáctví na výkon v této soutěži, ke kterým se nyní můžeme přesunout.

## Výsledky

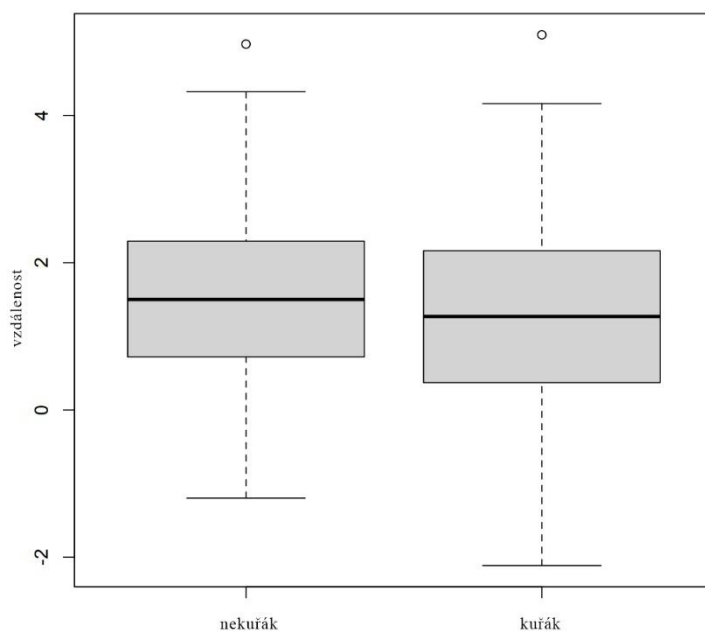
Po diagnostice samotného modelu jsme vyhodnotili vzdálenosti „dofouknutí“ peříčka zvláště pro kuřáky a nekuřáky, což můžete vidět v setu grafů 3. Na levém obrázku je zobrazen graf, který ukazuje distribuci závislé proměnné pro obě skupiny samostatně. V pravé části obrázku pak vidíme stejnou vizualizaci, avšak obě skupiny jsou zobrazeny v jedné tabulce s různě zbarvenými sloupci pro každou skupinu. Tyto grafy nám umožňují porovnávat distribuci výsledků mezi kuřáky a nekuřáky.

**Set grafů 3:** Distribuce vzdálenosti dofouknutí pro skupiny kuřáků a nekuřáků



Pro další vykreslení výsledků jsme také použili tzv. Box ploty, které lze vidět na obrázku 5, které také porovnávají výkon kuřáků a nekuřáků, lze na nich však vidět i rozptyl obou skupin.

**Graf 4:** Box ploty pro detailní vykreslení výsledků obou skupin



**Závěr:**

Naše analýza neprokázala statisticky významný rozdíl v dosaženém výkonu mezi kuřáky a nekuřáky. I když průměrný výkon nekuřáků byl vyšší než u kuřáků, rozdíl nebyl dostatečně velký na to, aby byl označen za statisticky signifikantní. Analýza zahrnovala výpočet intervalů spolehlivosti (CIs) pro různé úrovně spolehlivosti (90%, 95%, 99% a 99,9%). Na všech úrovních spolehlivosti se ukázalo, že se kuřáci a nekuřáci statisticky významně neliší v tom, jak daleko dokážou dofouknout peří svým dechem ( $p > 0,10$ ).

## Souhrn

Celkově naše analýza poskytuje důležité poznatky o vztahu mezi kouřením cigaret a výkonem ve výdechu. Především však ukazuje, jak efektivně lze využít smíšené modely pro analýzu dat s hierarchickou strukturou. Tyto poznatky mohou být prospěšné pro další výzkum v oblasti fyziologického výkonu a zdravotního chování, a mohou také posloužit jako inspirace pro další využití statistických metod v podobných studiích.

## ZDROJE:

Das, S. K. (2003). Harmful health effects of cigarette smoking. *Mol Cell Biochem* **253**, 159–165 (2003). <https://doi.org/10.1023/A:1026024829294>

Feschenko, Y. I., Illyinskaya, I. F., Arefieva, L. V., & Kuryk, L. M. (2018). Non-controlled bronchial asthma: the contemporary condition of the problem. *SO National Institute Physiology and Pulmonology named after F. G. Yanovsky NAMS of Ukraine*.

Jayes, L., Haslam, P. L., Gratziau, C. G., Powell, P., Britton, J., Vardavas, C., Jimenez-Ruiz, C., & Leonardi-Bee, J.; Tobacco Control Committee of the European Respiratory Society. (2016). SmokeHaz: Systematic Reviews and Meta-analyses of the Effects of Smoking on Respiratory Health. *Chest*, 150(1), 164-179. <https://doi.org/10.1016/j.chest.2016.03.060>

Sherman, C. B. (1991). Health Effects of Cigarette Smoking. *Clinics in Chest Medicine*, 12(4), 643-658. [https://doi.org/10.1016/S0272-5231\(21\)00814-5](https://doi.org/10.1016/S0272-5231(21)00814-5)