

Lineární regrese na různých vlastnostech písniček v závislosti na roce vydání

Adam Červenka

August 28, 2023

1 Úvod

Nejdříve jsem chtěl zkoumat změnu hlasitosti u písniček mezi lety 2000 a 2023 v závislosti na roce vydání. Ale při zkoumání dat jsem objevil 2 zásadní nedostatky: rozsah hlasitosti v datech byl velmi malý, pohyboval se od -16 do -1 a zároveň byl vždy celočíselný. Po tomto zjištění jsem tedy začal zkoumat statistické změny jiných vlastností u písniček v závislosti na roce vydání a zaujali mě konkrétní 2 vlastnosti: délka trvání a "valence", v překladu pozitivnost. U těchto 2 parametrů jsem pozoroval významné změny, nejlépe je to vidět na obrázku regrese.

2 O datasetu

Dataset pochází z Kaggle.com, uživatel "CONOR VAN EDEN" sesbíral data nejposlouchanějších písniček celosvětově mezi roky 2000 a 2023, k nim zahrnul i další jejich vlastnosti (jméno, rok, albm, dB, akustičnost, živost, pozitivnost atd., dohromady 13). Písniček je v datasetu 2385 (zhruba 100 za každý rok). Odkaz na dataset: [data](#)

2.1 Filtrování dat

Všiml jsem si, že v datech bylo pár písniček z let před rokem 2000, zároveň některé písničky byly delší než je obvyklé, proto jsem z dat vyhodil všechny písničky starší roku 2000 a delší než 400 sekund (6 minut 40 sekund, udělal jsem manuální kontrolu dlouhých písniček a vypadalo to na chybu v datasetu, dlouhé písničky ve skutečnosti nebyly dlouhé). Celkem se takto vyfiltrovalo 57 písniček, to je 2,4 % ze všech.

3 Výsledky

Níže jsou obrázky regrese, tečky odpovídají písničkám.

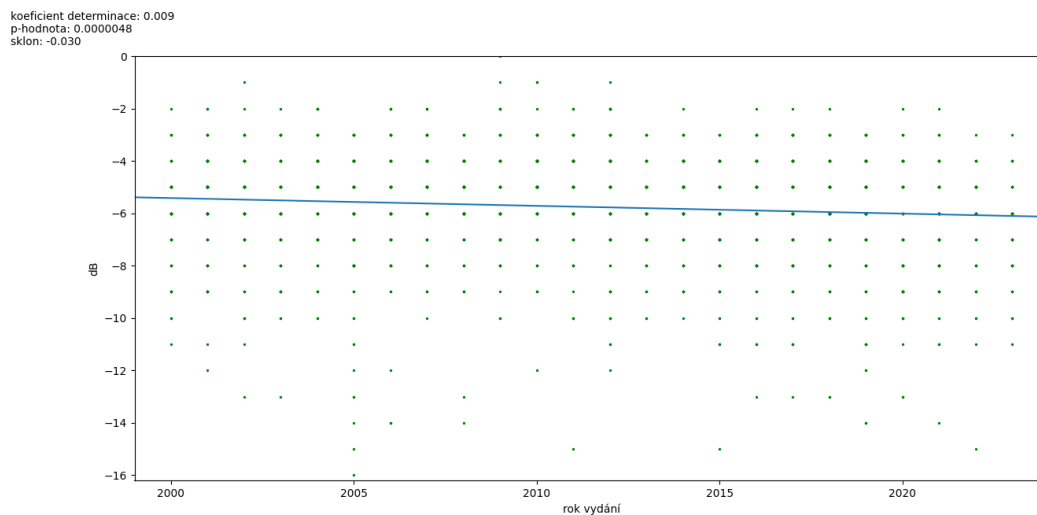


Figure 1: Lineární regrese hlasitosti v dB

P-hodnoty u všech 3 sledovaných ukazatelů jsou nízké, nižší než úroveň významnosti 0,05, takže víme, že veličiny na sobě jsou závislé. U všech sledovaných ukazatelů byla regrese klesající, přičemž u hlasitosti byla přímka velmi pomalu klesající, takže když vezmeme v potaz p-hodnotu a sklon přímky, tak statisticky se hlasitost písniček téměř nemění nebo se mění jen minimálně. U délky a pozitivnosti to ale bylo jinak. Začneme délkou.

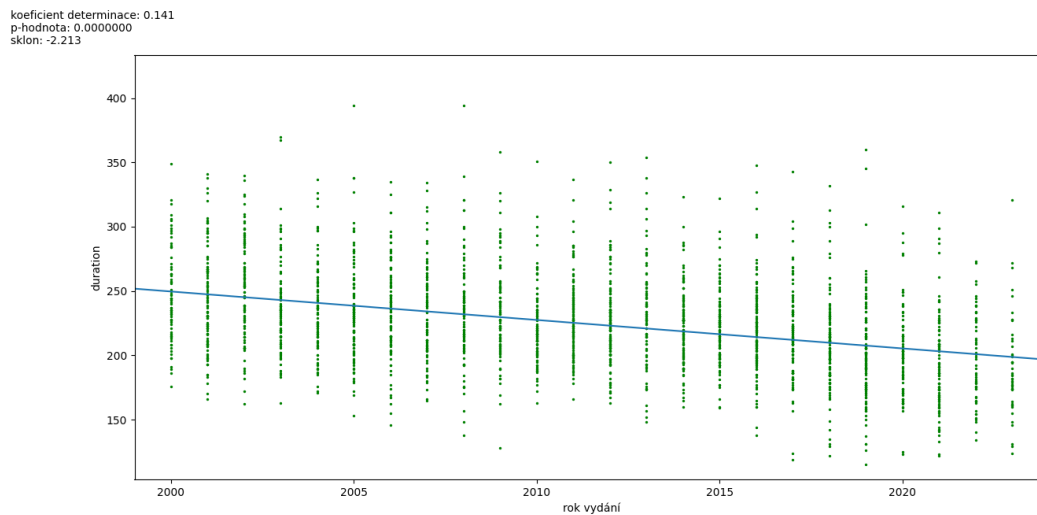


Figure 2: Lineární regrese délky trvání

Délka má výrazně klesající lineární regresi. Dokonce koeficient determinace není úplně nízký (14,5 %), je krásně vidět, že délka písniček má klesající tendenci z průměrných zhruba 250 sekund před 23 lety na zhruba 200 sekund dnes.

koeficient determinace: 0.053
p-hodnota: 0.0000000
sklon: -0.781

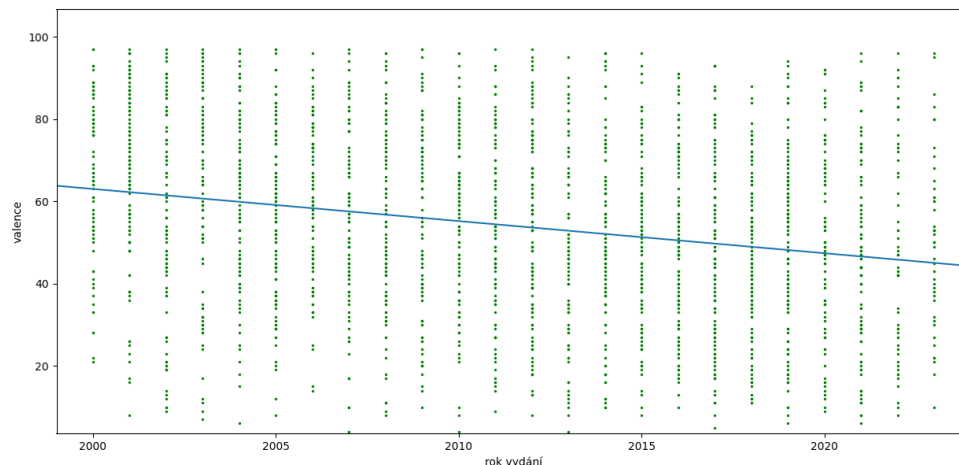


Figure 3: Lineární regrese pozitivnosti

Pozitivnost má také klesající lineární regresi, ne tak výrazně jako délka, ale pořád nezanedbatelně. Koeficient determinace je v tomto případě nízký, zřejmě kvůli velkému rozptylu hodnot pozitivnosti.

4 Závěr

Zjistil jsem, že hlasitost písniček je podobná nyní i před 23 lety. Dále jsem zjistil, že délka trvání písniček se zmenšuje trendem zhruba 2,2 sekundy za rok. A nakonec jsem zjistil, že pozitivnost je také klesající.

Možná vysvětlení:

Lidské ucho se za 23 let nezměnilo a moc ani technologie přehrávání zvuku, není třeba měnit hlasitost.

Komunikace mezi lidmi se zrychluje, zároveň kvůli různým krátkým návykovým videím lidé neudrží tak dlouho pozornost na jedno video/písničku, možná na to reagují autoři písniček pomocí zkrácení a zahuštění obsahu písniček.

Zdá se mi, že v okolí mají lidé často různé psychické problémy, nedokážu říct s čím to souvisí, ale pozitivnost tím klesá. Písničky jen odrážejí náladu ve společnosti. Zároveň poslední cca 3 roky nebyly pozitivní v žádném smyslu, měli jsme a máme tu ekonomické a sociální problémy.