

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

BAKALÁRSKA PRÁCA

Analýza športových dát relativnej povahy



Katedra matematické analýzy a aplikací matematiky

Vedoucí bakalářské práce: **prof. RNDr. Karel Hron, Ph.D.**

Vypracovala: **Dominika Kopecká**

Studijní program: B1103 Aplikovaná matematika

Studijní obor Matematika-ekonomie se zaměřením na bankovníctví/pojišťovnictví

Forma studia: prezenční

Rok odevzdání: 2022

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Dominika Kopecká

Název práce: Analýza športových dát relatívnej povahy

Typ práce: Bakalárska práca

Pracovišťa: Katedra matematickej analýzy a aplikácií matematiky

Vedoucí práce: prof. RNDr. Karel Hron, Ph.D.

Rok obhajoby práce: 2022

Abstrakt: Bakalárska práca sa zaoberá analýzou dát športovca, v tomto prípade cyklistickými dátami relatívnej povahy. V prvej časti sú predstavené silové zóny, základné informácie o tréningu a sú špecifikované typy tréningov a preteky. Bližšie je tu popísaný polarizovaný tréningový plán, ktorého efektívnosť je neskôr skúmaná. Dáta relatívnej povahy sú brané od jedného športovca a reprezentujú čas strávený v jednotlivých silových zónach počas jeho aktivity. Cieľom práce je pozrieť sa na relatívnu štruktúru tréningových aktivít a polarizovaného tréningového plánu. Na analýzu sú použité dve metódy prieskumovej mnohorozmernej štatistiky, a to analýza hlavných komponentov a lineárna regresia s ohľadom na kompozičné dáta, pričom ich výsledky sú ďalej podrobne interpretované.

Kľúčová slova: kompozičné dáta, cyklistické tréningy, lineárna regresia, analýza hlavných komponentov

Počet strán: 42

Počet príloh: 2

Jazyk: slovenský

BIBLIOGRAPHICAL IDENTIFICATION

Author: Dominika Kopecká

Title: Analysis of sports compositional data

Type of thesis: Bachelor's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: prof. RNDr. Karel Hron, Ph.D.

The year of presentation: 2022

Abstract: The topic of the bachelor thesis is an analysis of sports data, specifically cycling data of relative nature (compositional). The first part introduces the power zones, basic information about training and specifies different types of trainings and races. Polarized training plan is described in more detail, the effectiveness of which is explored later. Compositional data are taken from one cyclist and represent the time spent in each power zone during his activity. The aim of the thesis is to look at the relative structure of training activities and polarized training plan. Two methods of explanatory multivariate statistics are used for the analysis, namely the principal component analysis and linear regression with respect to the compositional data, and their results are further interpreted in detail.

Key words: compositional data, cycling trainings, linear regression, principal component analysis

Number of pages: 42

Number of appendices: 2

Language: Slovak

Prohlášení

Prehlasujem, že som bakalársku prácu spracovala samostatne po vedením pána prof. RNDr. Karla Hrona, Ph.D. a všetky použité zdroje som uviedla v zozname literatúry.

V Olomouci dňa

.....

podpis

Obsah

Úvod	7
1 Základy tréningovania	8
1.1 Merač výkonu	8
1.2 Funkčný prahový výkon	9
1.3 Normalizovaná sila a faktor intenzity	11
1.4 Skóre tréningového zaťaženia	12
1.5 Tréningové zóny	12
1.5.1 Popis jednotlivých zón	14
1.6 Tréningový plán	15
1.6.1 Tréningy	15
1.6.2 Polarizovaný tréningový plán	16
2 Kompozičné dáta	18
2.1 Súradnicové reprezentácie kompozičných dát	19
3 Vybrané štatistické metódy	22
3.1 Analýza hlavných komponentov	22
3.2 Lineárna regresia	23
3.2.1 Lineárna regresia s kvalitatívnymi premennými	24
3.2.2 Lineárna regresia kompozičných dát s reálnou vysvetľovanou premenou	25
4 Analýza dát	27
4.1 Predstavenie dát	27
4.1.1 Dáta na analyzovanie štruktúry typu aktivity	27
4.1.2 Dáta na analyzovanie polarizovaného tréningového plánu	29
4.2 PCA analýza	30
4.3 Lineárna regresia	34
4.4 Absolútny rozdiel v porovnávaní dvoch období	39
Záver	40
Literatura	42

Poděkování

Ďakujem môjmu vedúcemu bakalárskej práce prof. RNDr. Karlovi Hronovi, Ph.D. za odbornú pomoc a užitočné rady, no tiež za ľudský prístup, ochotu a porozumenie. Taktiež ďakujem cyklistovi Przemyslawovi Kotulskému za poskytnutie dát potrebných pre praktickú časť mojej práce.

Úvod

Témou mojej bakalárskej práce je analýza športových dát relatívnej povahy, konkrétne som sa zamerala na jednotlivé športové činnosti cyklistu ako sú tréningy a preteky. Nie je novinkou, že matematika sa v športe využíva už odpradáva. Výnimkou nie je ani cyklistika. Vyskytuje sa všade. Či už hovoríme o mierach bicykla, minimalizovania odporu vzduchu alebo o jednoduchých výpočtoch trás. Svoje uplatnenie si našla aj v tréningoch, kde je s obľubou využívaná množstvom profesionálnych trénerov. Spracovávanie dát a postupné prispôbovanie tréningu je kľúčom k úspechu cyklistu. Zachovanie štruktúry tréningu, dodržiavanie presne vypočítaných hodnôt sily, ktorú je potrebné vydať v presnom čase, je len zlomok toho, čo sa za tréningami reálne deje. Existuje mnoho rôznych prístupov k trénovaniu a maximalizovaniu sily cyklistu. V tejto práci sa chcem pozrieť hlbšie do toho, čo sa skrýva za niečím tak jednoducho vyzerajúcim, ako je obyčajná jazda na bicykli. Mojim cieľom je aplikovať dve vybrané štatistické metódy a sledovať relatívnu štruktúru rôznych typov tréningov. Taktiež sa chcem pozrieť na to, či pomáha štruktúrovaný tréningový plán na zlepšenie výkonnosti. Zaujíma ma tiež, či možno už z malého vzorku od jedného cyklistu odvodiť plnohodnotný záver o tréningoch. V mojej práci sa primárne zameriavam na dáta relatívnej povahy, čo nepatrí k mainstreamovému prístupu, takže verím, že aj z tohto pohľadu bude moja práca pre čitateľov obohacujúca.

Kapitola 1

Základy trénovanía

Na to, aby moja práca poskytla čitateľovi komplexný obraz o trénovaní a pochopenie jednotlivých pojmov a veličín, ktoré sú v analýze zahrnuté, je potrebné na začiatok vysvetliť pár pojmov a veličín. V tejto kapitole som čerpala zo zdrojov [4] a [1].

1.1. Merač výkonu

Medzi cyklistami je stále viac a viac populárny merač výkonu, takzvaný power meter. Pri cyklistoch na špičkovej úrovni je to základom ich športovej výbavy. Je to zariadenie pripevnené na bicykli, ktoré meria ich výkon. Dôvodom prečo je taký obľúbený a vyhľadávaný je to, že športovci sú schopní kvantitatívne sledovať úroveň ich kondície, jednoduchšie definovať ich slabé stránky a na základe toho zostaviť optimálny tréning. Merač výkonu nahráva enormné množstvo dát, ktoré je následne možné stiahnuť. Vďaka tomu je umožnená dôsledná analýza každej časti tréningu, každého kopca, šprintu a toho, či jazdec splnil svoje ciele alebo musí prehodnotiť svoje tréningové metódy. Predtým ako boli vynájdené merače výkonu, cyklisti len odhadovali miesta, v ktorých boli silní a slabí, čo často krát viedlo k chybám a miesto zlepšenia, prichádzalo iba zhoršenie. S vynájdenním power metra prešiel cyklistický svet revolúciou a už po pár meraniach na pretekoch a tréningoch bolo možné určiť s veľkou presnosťou, kde sa cyklista potrebuje zlepšiť a naopak, kde má výhodu nad ostatnými. Spracovávať dáta si môže každý

cyklista alebo jeho tréner sám, dnes však na to existuje už aj veľa športových softvérov, ktoré to robia za nich.

1.2. Funkčný prahový výkon

Funkčný prahový výkon (*ang. Functional threshold power, FTP*) ako už názov napovedá, je určitý prah, teda hranica v tele športovca, resp. cyklistu. Je to najvyšší možný výkon udávaný vo wattoch, ktorý je schopný cyklista udržať kvázi neobmedzene dlho bez toho, aby sa unavil. Pointou toho je, že ak cyklista prekročí svoju hodnotu FTP, unaví sa omnoho skôr (v praxi u dobre trénovaných cyklistov je to približne po jednej hodine), zatiaľčo ak bude jeho sila tesne pod hodnotou FTP, vydrží jazdiť značne dlhšie. Že sa unaví znamená, že presiahne svoj anaeróbny prah (nazývaný tiež laktátový prah), tzn. takú intenzitu cvičenia, kedy sa začína zvyšovať množstvo laktátu v krvi. To je spôsobené tým, že pri nižších intenzitách telo spaľuje tuky, zatiaľčo pri vyšších sa začínajú spaľovať aj cukry. Ak sa teda cyklista dostane na svoj anaeróbny prah a pokračuje ďalej, svalom ubúda kyslík a teda, ak už ho nemôžu využiť, získavajú energiu zo štiepenia jednoduchých cukrov. To je však spojené s tvorbou laktátu, teda soli kyseliny mliečnej. Navonok sa to prejavuje vysokým tepom, zadýchaním, únavou a pocitom ubúdania energie. Ak je cyklista za týmto prahom už nejakú dobu (zväčša vydrží len pár minút, trénovaní cyklisti aj trochu dlhšie), množstvo laktátu vo svaloch znižuje pH a tak sa zakyslia, teda vznikne svalová stuhnutosť a bolesť, čo v praxi dokáže nohy takmer zastaviť. Poznať hodnotu funkčného prahového výkonu je pre cyklistu nielen dôležité, ale aj potrebné, ak sa chce zlepšovať a vidieť pokrok. Je to taktiež jediná hodnota, ktorá reálne ukazuje zlepšenie alebo zhoršenie, nakoľko nie je ovplyvnená počasím alebo profilom terénu, iba silou vydanou cyklistovými nohami. Je veľmi dôležité si však uvedomiť to, že hodnota FTP nie je na porovnávanie cyklistov medzi sebou, ale na porovnávanie výkonu jedného cyklistu v čase.

Príklad 1. Cyklista 1 má 85 kg a FTP 300 W. Cyklista 2 má 61 kg a FTP 260 W. Teda Cyklista 1 má omnoho väčšie FTP a teda by mal byť lepší. To je však irelevantné, pretože FTP je závislé na váhe. Preto sa často prepočítava na jednotky watty na kilogram. V našom prípade Cyklista 1 má približne hodnotu 3,529 W/kg a Cyklista 2 má 4,26 W/kg. Z toho jasne vidno, že je na tom

lepšie Cyklista 2. Avšak to stále nie je dobrý dôvod na to, aby bolo porovnávanie medzi jazdcami relevantné. Hlavnou príčinou je to, že aj napriek tomu, že Cyklista 2 je zdatnejší, na rovine má omnoho väčšiu výhodu Cyklista 1 vďaka jeho váhe, ktorá mu zaručuje lepšiu zotrvačnosť. Avšak ak by bol Cyklista 1 hobby a Cyklista 2 trénovaný profesionál, v takom prípade by bol rýchlejší Cyklista 2 aj napriek nižšej váhe. Z toho je jasné, že FTP hodnota je len ukazovateľom zlepšenia resp. zhoršenia cyklistu ako jedinca. Na to, aby cyklista vedel na čom je, existujú všeobecné tabuľky s rôznymi hodnotami W/kg. Nazýva sa Power Profile Chart a je rozdelená na 9 kategórií s rôznymi hodnotami W/kg. Okrem hodnôt FTP sa tam nachádzajú aj hodnoty prepočítané na W/kg, ktoré ukazujú maximálnu možnú silu vydanú za 5 sekúnd, 1 minútu, 5 minút a to ako pre mužov, tak aj pre ženy.

Ako získať hodnotu FTP

FTP sa získava tak, že cyklista vydá zo seba maximálnu silu po dobu jednej hodiny a priemerná sila (*ang. Average power, AWP*) je jeho hodnota FTP. Avšak v praxi by bol takýto test možný iba na internom trenáži, keďže podmienky vonku, ako napríklad križovatky, semaforey, zvlnená cesta, prekážky na ceste a podobne, neumožňujú ísť hodinu naplno. Ekvivalentný k hodinovému testu je aj 20 minútový test. Teda cyklista musí ísť 20 minút maximálny výkon a potom jeho AWP je potrebné prenásobiť faktorom 0,95. Teda 95% z maximálneho výkonu na 20 minút je rovný hodnote FTP cyklistu. Takýto test už je možný ako von, tak aj dnu. Reálne však znova trvá okolo hodiny, keďže cyklista by pred začatím 20 minútového maxima mal rozohriať svaly, aby sa telo pripravilo na záťaž a rov-

nako aj potom by mal nohy vyjazdiť a upokojiť organizmus po stresujúcom zážitku.

1.3. Normalizovaná sila a faktor intenzity

Bicyklovanie, tréovanie a pretekanie má veľa rôznych variabilných veličín, všeobecne je to cvičenie so stochastickým charakterom. Priemerná sila za tréning často nestačí, pretože cyklista často ide dole kopcom alebo ho tlačí dopredu zotrvačná sila. V tejto časti netlačí do pedálov a respektíve neprodukuje žiadnu silu. Na to je dobrá normalizovaná sila (*ang. Normalized power, NP*), ktorá tento fakt zohľadňuje. Z tréningu berie dáta ako zrýchlenie, náročnosť, vzniknutý stres, záťaž, ale aj využívanie glykogénu a produkciu laktátu. Všetky tieto dáta spracováva softvér na to určený a výsledná hodnota je teda sila vo wattoch, ktorú by bol schopný cyklista vyprodukovať za rovnakú cenu unavenia sa, ak by jeho výsledok bol perfektne konštantný.

Príklad 2. Softvér vyrátal cyklistovi z 3 hodinového tréningu NP 180 W, pričom jeho priemerná sila bola len 130 W. Priemerná sila, ktorú by bol schopný vyprodukovať na trénažeri za dokonalých podmienok za 3 hodiny, s tým že by sa unavil rovnako, by bola práve 180 W.

Normalizovaná sila je už omnoho presnejším ukazovateľom ako priemerná sila, ale faktor intenzity (*ang. Intensity factor, IF*) ešte berie do úvahy aj športovcov individuálny priebeh výkonu v čase a jednoducho sa vyrátava ako podiel normalizovanej sily a FTP,

$$IF = \frac{NP}{FTP}.$$

1.4. Skóre tréningového zaťaženia

V tréningoch je veľmi dôležitá intenzita, ale aj dĺžka stráveného času. Obe veličiny zvyšujú náročnosť tréningu. Avšak je treba dôležité si uvedomiť, že sú na sebe závislé. Táto závislosť je vyjadrená ako skóre tréningového zaťaženia (*ang. Training stress score, TSS*), ktoré zohľadňuje obe veličiny a kvantitatívne ukazuje hodnotu celkového zaťaženia tréningu. Jednoducho sa dá vyrátať z každého tréningového dňa, pričom dnes už to robia často aj športové softvéry.

$$TSS = \frac{s \times W \times IF}{FTP \times 3600} \times 100,$$

kde s je trvanie tréningu v sekundách, W je normalizovaná sila vo wattoch a číslo 3600 označuje počet sekúnd v jednej hodine.

TSS si vieme predstaviť aj ako hodnotu, ktorá približne ukazuje, koľko glykogénu telo využilo v každom tréningu. Pointou výpočtu je to, že TSS sa vyrátava akoby z jedno-hodinovej časovky na maximum. Teda športovec, ktorý ide hodinu na svojom FTP, by mal skóre 100 TSS bodov, pričom faktor intenzity by bol 1.0. Aj v tomto prípade je veľmi dôležité si uvedomiť, že táto hodnota je závislá na športovcovi. Teda rovnaký tréning môže mať inú hodnotu TSS pri rôznych cyklistoch. TSS je dôležité na to, aby bolo možné určiť unavenie tela a následne zvoliť primeraný oddych na zotavenie.

TSS	Intenzita	Status zotavenia
<150	Nízka	Telo je zotavené na druhý deň
150 - 300	Mierna	Na druhý deň je možné cítiť jemnú únavu, ale všeobecne by malo byť telo zotavené na druhý deň
300 - 450	Vysoká	Je možné cítiť jemnú únavu po dvoch dňoch
>450	Veľmi vysoká	Je pravdepodobná únava počas pár dní

Obrázek 1.1: Vplyv tréningového stresu na únavu

1.5. Tréningové zóny

Tréningové zóny sú zóny každého cyklistu založené na jeho pocitoch záťaže, sile alebo tepe. Využívajú sa na prispôbenie tréningu, podľa toho, čo je potrebné zlepšiť. Tieto zóny má každý cyklista individuálne a prepočítané na svoje

fyziologické zdatnosti. Sú základom dobrého tréningu a sú využívané vo vrcholovej cyklistike najlepšimi trénermi. Na túto kapitolu som použila zdroje [4] a [1].

Základné tréningové zóny vytvoril celosvetovo uznávaný cvičebný fyziológ a majster sveta v časovke Andrew Coggan. Prvým a najstarším typom zón je takzvaná Borgova stupnica hodnotenia vnímanej záťaže (*ang. Rating of perceived exertion, RPE*), ktorú si jazdec volí sám na základe vlastného vnímania pocitu záťaže. Skúsení cyklisti to dokážu odhadnúť pomerne presne aj bez merača

Hodnotenie	Popis záťaže
0	Žiadna
0.5	Extrémne nízka
1	Veľmi nízka
2	Ľahká
3	Mierna
4	Trochu ťažká
5	Ťažká
6 - 7	Veľmi ťažká
8 - 9	Extrémne ťažká
10	Maximálna

Obrázek 1.2: Borgova stupnica hodnotenia vnímanej záťaže

výkonu. Tým sa dostávame k siedmym klasickým zónam od doktora Coggana, ktoré sú vypočítané z FTP daného cyklistu. Podľa neho je 7 zón minimum na to, aby sme presne dokázali popísať štruktúru tréningu. V praxi pri rôznych typoch tréningoch však sú využívané aj iné škály, napríklad 3 zóny alebo 6 zón.

Zóna	Názov	% z FTP	% z FTTH	RPE	Typická dĺžka jazdy (min)	Typická dĺžka intervalu (min)
1	Regenerácia	<55	<68	<2	30 - 90	N/A
2	Aerobný	56 - 75	69 - 83	2 - 3	60 - 300	N/A
3	Tempo	76 - 90	84 - 94	3 - 4	60 - 180	N/A
4	Prahový	91 - 105	95 - 105	4 - 5	N/A	8 - 30
5	Aerobná kapacita	106 - 120	>106	6 - 7	N/A	3 - 8
6	Anaerobná kapacita	121 - 150	N/A	>7	N/A	0,5 - 3
7	Maximálna sila	N/A	N/A	maximum	N/A	< 0,5

Obrázek 1.3: Klasické tréningové zóny

1.5.1. Popis jednotlivých zón

Pre prvú zónu je charakteristické pomalé dýchanie, iba jemne zvýšený tep a ľahké tlačenie do pedálov. Jazda by mala byť pomalá a pokojná. Tento typ sa zväčša využíva po pretekoch, po ťažkom tréningu a intervaloch na to, aby sa telo zotavilo.

Pri druhej zóne by sa malo dýchanie a tep pomaly zrýchľovať, jazda by však stále mala byť pokojná a cyklista by mal byť počas nej schopný rozprávať. Nevyužíva sa už na regeneráciu, ale na zlepšenie vytrvalosti.

Tretia zóna si už vyžaduje väčšiu koncentráciu, dýcha sa ťažšie, konverzácia počas jazdy začína byť nemožná.

V štvrtej zóne už je dýchanie ťažké a tep vysoký, je potrebná veľká telesná sila. Je ťažko sa tu udržať dlhší čas, preto je rozdelená na intervaly. Cyklista sa začína cítiť nepohodlne a začínajú bolesti nôh.

Piata zóna je bolestivá a rýchla. Cyklista potrebuje nielen veľkú silu tela, ale veľkú rolu tu zohráva aj mentálna sila. Je cítiť značné pálenie v nohách a nie je možné udržať túto zónu viac ako pár minút vkuse.

Šiesta zóna má bolieť. Tep má byť na maxime. Veľmi ťažké dýchanie. Túto zónu by mal cyklista zvládnuť naraz len okolo 30 sekúnd.

Siedma zóna je maximum, ktoré zo seba dokáže cyklista vydať, inak nazývané aj maximálny sprint. Zväčša trvá len pár sekúnd, preto sa často v modeloch nevyužíva alebo sa spája so šiestou zónou.

Individualizované tréningové zóny

Všade sa nájdu výnimky a tak to je aj v cyklistike. Veľa jazdcov nespadá do modelu klasických silových zón a vznikajú outliery. To znamená, že aj napriek tomu, že zóny sú vyrátané priamo na osobu, môže sa stať, že v zóne v ktorej by nemal dlho vydržať, vydrží omnoho dlhšie a naopak. Aby sa znížil počet takýchto výnimiek a aby bol model viac aplikovateľnejší, pridali sa medzi zóny ďalšie dve, pričom jedna z nich je nazývaná Sweet spot a je zásadnou pri tréningoch.

1.6. Tréningový plán

V tejto kapitole som čerpala zo zdroja [4]. Tréning je štruktúrovaný celoročný plán športovca. Nazýva sa to ročný tréningový plán a je rozdelený na niekoľko častí. Takzvané makrocikly (zväčša sa používa jeden makrociklus – tréningový rok, ale môžu sa vyskytnúť aj dva a viac), mezocikly (stredne dlhé obdobia, zvyčajne pár týždňov), mikrocikly (obvykle trvá týždeň) a tréningové jednotky. Tréning spočíva v periodizácii mezocyklov v závislosti od toho, čo je potrebné zlepšovať.

Mezocikly

Najdôležitejšie je správne rozdeliť rok do mezocyklov. Máme rôzne typy, napríklad na zvýšenie výkonnosti, lepšiu fyzickú kondíciu, techniku a podobne. V praxi trvá jeden mezociklus od 2 do 8 týždňov. Počas roka sa ich strieda viacero. V závislosti od času v roku, vyberáme jeho typ. Môže to byť prípravné obdobie, vrcholové alebo prechodné.

1.6.1. Tréningy

V mojej bakalárskej práci je dôležité vysvetliť jednotlivé typy tréningov. Ako je spomínané vyššie, každý mezociklus má nejaký cieľ. K tomuto cieľu je potrebné sa dostať zostavením rôznych tréningov v presnej postupnosti za sebou. Sú stavané taktiež na to, aby mal športovec najvyššiu formu na pretekoch a aby bol čo najviac oddýchnutý. Ďalej sa budem sústreďovať na dva typy pretekov, a to dlhé cestné preteky (*ang. Road race*), ktoré zvyčajne trvajú pár hodín a majú okolo 70 - 200 km a kritéria (*ang. Kriteriums*), ktoré sú špecifické tým, že sa jazdia kolá na rovine vo veľkej rýchlosti, väčšinou majú do 100 km a bodujú sa priebežne, čo znamená, že sa môže bodovať každé piate alebo tretie kolo, s tým, že kto ho prvý prejde, dostane najviac bodov. Pre Road Race je typický veľký výkon a rovnomerne rozložený čas vo všetkých silových zónach, no dominujú zóny vyššej intenzity, pričom pri kritériách je to vysoký výkon počas bodovacích kôl a nižší počas ostatných, preto sa jazdí aj v nižších zónach na menšej intenzite.

Tréningy delíme podľa nasledujúcich typov, ktoré potom budeme ďalej uvažovať:

Zotavovací tréning (*ang. Recovery training*) - je to pomalá jazda, kde sa telo zotavuje z ťažkého tréningu alebo pretekov, jazdí sa väčšinou v prvých dvoch zónach.

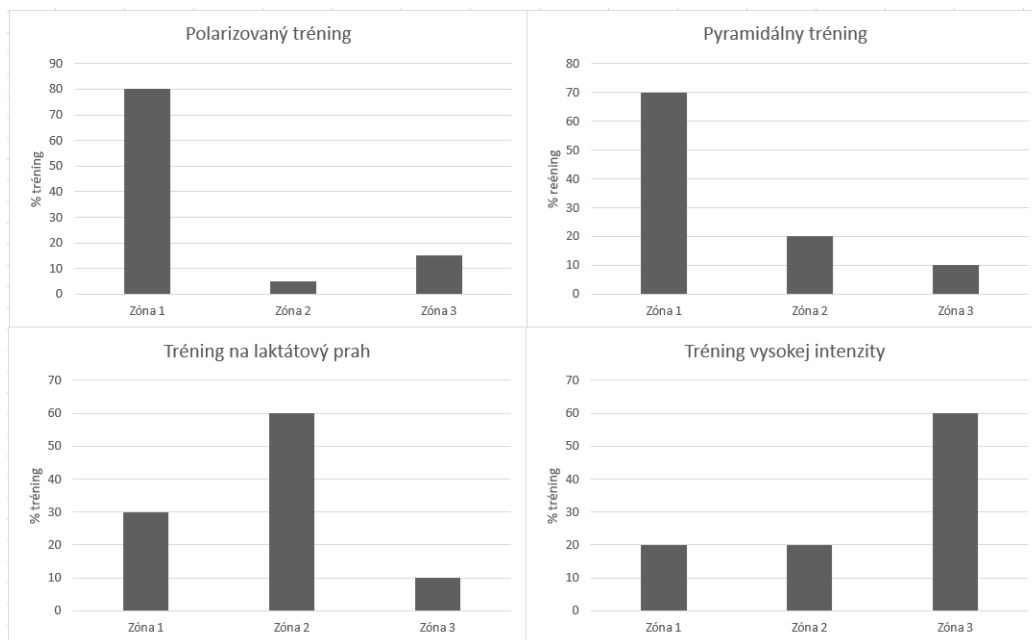
Ľahký tréning (*ang. Easy ride*) - tréning v nízkej intenzite, medzi recovery a endurance, dominujú tu nižšie a stredné zóny.

Vytrvalostný tréning (*ang. Endurance training*) - ťažší tréning, väčšinou dlhý, pripravuje telo na záťaž, najviac by sa mal jazdiť v stredných a vyšších zónach.

Intervaly (*ang. Intervals*) - ťažký tréning, môže sa trénovať ako v nižších, tak aj v stredných zónach, no obsahuje úseky vo vysokých zónach, môžu mať opakovania po jednej minúte alebo po piatich, desiatich alebo viacerých minútach, je veľmi individuálny a dá sa rôzne kombinovať.

1.6.2. Polarizovaný tréningový plán

Existuje mnoho spôsobov, akým sa dajú tréningy poskladať. Sú modely, podľa ktorých sa tréneri riadia a skladajú mezocykly. Ja som sa v mojej práci zamerala na polarizovaný tréning. Vychádza z toho, že cyklisti trávia veľa času v stredných zónach. Teda telo si neodpočinie, ale zároveň ani netrénuje svoje maximum. V mnohých prípadoch telo nedokáže maximum ani dosiahnuť, práve kvôli tomu, že je unavené. V tom spočíva pointa polarizovaného tréningu. Na jeho vysvetlenie je v prvom rade potrebné zmieniť, že ide o tréning, ktorý sa trénuje iba v 3 zónach. Prvá zóna pritom označuje základný tréning pod aerobickým prahom, druhá tréning vo vyššej intenzite alebo tzv. sweet spot a tretia sú vysokointenzívne tréningy. Pri polarizovanom tréningu ide o najväčšiu možnú minimalizáciu trávenia času v druhej zóne. 80% tréningov by malo byť v prvej zóne, takmer žiaden v druhej a 15-20% v tretej zóne. V prepočítaní na čas strávený v jednotlivých zónach, by to malo byť 90% v prvej, málo v druhej a okolo 10% v tretej, pričom sa to dá takiež upraviť aj na 95:5. Tieto tri zóny sa dajú vypočítať rôznymi metódami, pričom jedna z nich je napríklad aj z FTP.



Obrázek 1.4: Čas strávený v jednotlivých zónach - porovnanie polarizovaného tréningu s ostatnými druhmi tréningov

Kapitola 2

Kompozičné dáta

Kompozičné dáta si jednoducho vieme predstaviť ako taký typ štatistických dát, ktoré tvoria určitú časť celku a nesú relatívnu informáciu; typicky sú reprezentované ako dáta, ktorých súčet zložiek sa sčítava na konštantu. V praxi sú najjednoduchšie predstaviteľné percentá alebo zlomky, ktorých súčet sa rovná 1. Kompozičné dáta sa väčšinou sčítavajú na rovnakú konštantu, no sú prípady, kedy sa konštanta pri jednotlivých pozorovaniach mení, stále sa však zachovávajú pomery medzi jednotlivými časťami a teda svoju relatívnu informáciu si dáta zachovávajú. A to je hlavnou úlohou kompozičných dát, nie sú relatívnu informáciu. Samozrejme záleží, čo skúmame a občas je potrebná aj absolútna informácia. Teda dáta môžu byť kompozičné, ale aj nemusia, záleží na tom, čo potrebujeme. V našom prípade porovnávame čas cyklistu strávený v jednotlivých zónach pri rôznych záťažových aktivitách. Zrejme je nám veľmi potrebná relatívna informácia, ale svoje využitie by si tu našla aj ta absolútna. Napríklad v prípade, že by sme chceli skúmať dĺžku stráveného času jednotlivkej aktivity cyklistu.

V týchto dátach by sa taktiež nemali objavovať nulové hodnoty. V prípadoch, kedy je to nevyhnutné, sa menia na veľmi malé číslo blízko nule, kde sa následne transformujú ako nenulové hodnoty transformáciami popísanými ďalej v texte. V našom prípade však skúmame iba aktivity cyklistu, kde sa v každej fáze vyskytuje aspoň pár sekúnd. Sú to aktivity s vyššou záťažovou náročnosťou. Na nulový čas strávený v 5. alebo 6. zóne by sme mohli naraziť len vtedy, že by aktivita cyklistu bola na nízkej úrovni, napríklad vyššie spomínaná recovery ride, čo zna-

mená, že telo si po pretekoch musí oddýchnuť. V prípade, kde pozorujeme čas strávený v zónach po týždňoch, sa nula taktiež nevyskytuje. V tejto kapitole som vychádzala zo zdrojov [3, 2].

John Aitchison v osemdesiatych rokoch minulého storočia ako prvý matematicky definoval kompozičné dáta. Najjednoduchším typom je kompozícia vyjadrená ako stĺpcový vektor $\mathbf{u} = (u_1, u_2, \dots, u_D)$, ktorý má D zložiek alebo častí celku $u_i, i = 1, \dots, D$, pre ktoré platí $u_1 + \dots + u_D = 1, u_i > 0$. Pre takto definovanú kompozíciu je vhodným výberovým priestorom jednotkový simplex

$$S^D = \{u_i > 0, \sum_{i=1}^D u_i = 1\}.$$

Zovšeobecnene teda hovoríme o D zložkových kompozíciách, ktoré sú stĺpcovými vektormi $\mathbf{x} = (x_1, x_2, \dots, x_D)$, pričom ich jednotlivé zložky nadobúdajú kladné hodnoty a nesú relatívnu informáciu, teda súčet ich zložiek je pre vlastnú štatistickú analýzu irelevantný.

2.1. Súradnicové reprezentácie kompozičných dát

Ďalšou dôležitou vlastnosťou dát relatívnej povahy, ktorá vychádza priamo z predchádzajúcich poznatkov je fakt, že jednotlivé zložky je možné vyjadriť pomocou ostatných zložiek. Jednoducho povedané, zložky sú na sebe závislé. Aj keď je to ďalšia vlastnosť, ktorá nám špecifikuje dáta, nie vždy je žiadúca. Dáta nespĺňajú predpoklady nezávislosti zložiek, či vôbec predpoklad o reálnom výberovom priestore mnohých štatistických metód, a teda v prípade, že by boli na ne použité, môže dôjsť ku chybným výsledkom.

Je dôležité spomenúť tzv. Aitchisonovu geometriu, algebraicko-geometrickú štruktúru, ktorá dokáže odrážať ako dimenziu kompozícií, tak aj mierkovú invariantnosť dát, t.j. informácia obsiahnutá v dátach sa nemení, ak ich prenásobíme ľubovoľnou kladnou konštantou. Prelomovým bodom bolo zavedenie súradníc v ortonormálnej báze Aitchisonovej geometrie. Vďaka nim bolo umožnené zobrazenie medzi Aitchisonovou geometriou a euklidovskou geometriou v reálnom priestore, pričom ostala zachovaná dimenzia a metrické vlastnosti kompozície.

Aitchison našiel spôsob, ako zanechať kompozičným dátam svoju relatívnu informáciu, ale aby na ne zároveň bolo možné použiť známe štatistické metódy. Zavedol logpodielovú metodiku. Jej cieľom bolo zaviesť transformácie dát na nové premenné tak, aby sa dáta transformovali z pôvodného výberového priestoru, simplexu, do reálneho priestoru. Následne by ich analyzovanie bolo jednoduchšie. Tieto transformácie sú založené na logaritmických podieloch.

Prvá transformácia označovaná *alr* je založená na aditívnych podieloch (*ang. additive logratio*) a druhá *clr* na centrovaných logaritmických podieloch (*ang. centered logratio*). Obe transformujú dáta zo simplexu do reálneho vektorového priestoru. Transformácia vektora \mathbf{x} na aditívnych podieloch je definovaná

$$alr(\mathbf{x}) = \left(\ln \frac{x_1}{x_D}, \ln \frac{x_2}{x_D}, \dots, \ln \frac{x_{D-1}}{x_D} \right),$$

kde D je počet zložiek kompozície \mathbf{x} . Pri tejto transformácii teda využívame logaritmus podielu, v ktorom delíme jednotlivé zložky jednou vybranou zložkou. V definícii je použitý posledný prvok ako deliteľ, ale v praxi sa samozrejme zložky vektora menia. A práve tento faktor je rozhodujúci pri tom, či je *alr* transformácia vhodná. Pri zmene poradia zložiek vo vektore sa menia aj výsledky analýz, pretože táto transformácia závisí na výbere deliteľa, čo určite nie je v konečnom dôsledku žiadúce.

Transformácia vektora \mathbf{x} na centrovaných podieloch je definovaná

$$clr(\mathbf{x}) = \left(\ln \frac{x_1}{g(\mathbf{x})}, \ln \frac{x_2}{g(\mathbf{x})}, \dots, \ln \frac{x_D}{g(\mathbf{x})} \right),$$

kde D je počet zložiek vektora \mathbf{x} a $g(\mathbf{x})$ je geometrický priemer zložiek vektora \mathbf{x} . Keďže pri tejto metóde je deliteľom geometrický podiel, prirodzene to rieši problém *alr* transformácie. Zachováva metrické vzdialenosti, ale vo výsledku vždy dáva singulárnu variančnú maticu, čo v mnohých prípadoch taktiež nie je výhodou.

Clr a *alr* transformácie vedú z geometrického hľadiska na súradnicový systém vzhľadom na generujúci systém, resp. bázu, ktorá nie je ortonormálna. Orto-

normálne súradnice dostaneme využitím izometrickej logpodielovej (*ang. izometric logratio*) transformácie definovanej ako

$$\mathbf{x} = (x_1, x_2, \dots, x_D) \rightarrow \mathbf{x}^{(l)} = (x_l^{(l)}, x_1^{(l)}, \dots, x_{l-1}^{(l)}, x_{l+1}^{(l)}, \dots, x_D^{(l)})$$

$$z_i^{(l)} = \sqrt{\frac{D-i}{D-i+1}} \ln \frac{x_1^{(l)}}{\sqrt[D-i]{\prod_{j=i+1}^D x_j^{(l)}}} \text{ pre } i = 1, \dots, D-1$$

pričom v tomto prípade, prvé súradnice obsahujú všetkú relatívnu informáciu (agregujú všetky príslušné logpodielky) s i -tou zložkou. Ostatné súradnice nám vysvetľujú zbytok informácie obsadenej v kompozícii.

Kapitola 3

Vybrané štatistické metódy

Pre úvodnú analýzu cyklistických dát, ktorá je predstavená v tejto bakalárskej práci, boli vybrané analýza hlavných komponentov a lineárna regresia - samozrejme s ohľadom na to, že pracujem s kompozičnými dátami.

3.1. Analýza hlavných komponentov

V tejto kapitole som použila zdroj [3]. Analýza hlavných komponentov (*ang. principal component analysis - PCA*) je štatistická metóda, ktorej cieľom je zredukovať dimenziu množiny vstupných dát vytvorením nových premenných (hlavných komponentov) tak, aby už niekoľko prvých z nich zachovalo čo najväčšie množstvo informácií o pôvodnej množine. Nové súradnice sa nazývajú hlavné komponenty. Ich počet je menší alebo rovný počtu pôvodných premenných. Informácie o danom súbore sú zachované vďaka tomu, že sa sleduje celkový rozptyl, ktorý je práve zdrojom informácií v dátach. Analýza hlavných komponentov redukuje dimenziu pôvodných údajov na dve (prípadne tri, štyri alebo viac) hlavné komponenty, ktoré sa dajú, v prípade dvoch, graficky znázorniť do biplotu. Ich počet závisí od pôvodného súboru a ich schopnosti zahrnúť čo najviac informácie o ňom. Biplot nám teda umožňuje v jednom grafe vidieť informácie ako o premenných, tak aj o jednotlivých pozorovaniach prostredníctvom záťažie (váhy pôvodných premenných v lineárnej kombinácii definujúca príslušný hlavný komponent), resp. skóru (súradnice dát v novom súradnicovom systéme tvoreným

hlavnými komponentami).

V prípade kompozičných dát sa PCA vykonáva v clr koeficientoch. Vzhľadom k ich singularnej variančnej matici sa pri interpretácii sústredíme najmä na vzdialenosti medzi vrcholmi šípok (reprezentujúce záťaž), ktoré aproximujú rozptyly príslušných párových logpodielov, a teda možnú preporcionalitu zložiek. V smere jednotlivých šípok sú potom umiestnené pozorovania (v biplote ako body), ktoré dominujú príslušnej kompozičnej zložke.

3.2. Lineárna regresia

V tejto kapitole bol použitý zdroj [5]. Lineárna regresia je jedna z najviac používaných metód regresnej analýzy. Regresná analýza popisuje vzťah medzi dvoma alebo viacerými kvantitatívnymi premennými, pričom lineárna regresia medzi nimi popisuje vzťah lineárny. Ten zapíšeme ako

$$Y = \beta_0 + \beta_1 X + \epsilon.$$

V tomto prípade ide o jednoduchú regresiu, kde Y je závislá premenná, ktorú odhadujeme a X je vysvetľujúca (nenáhodná) premenná. Parametre β_0 a β_1 sú neznáme regresné koeficienty modelu a ϵ je náhodná veličina zvaná náhodná chyba, pričom jej stredná hodnota je nulová a má konštantný nezáporný rozptyl.

Viacnásobná lineárna regresia odhaduje hodnoty premennej Y viacerými vysvetľujúcimi premennými X_1, X_2, \dots, X_p a jej regresný model má formu

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon,$$

a pre n -ticu pozorovaní nezávislých premenných a závislej premennej $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i)$ dostaneme

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i \text{ pre } i = 1, \dots, n.$$

Keďže sú parametre $\beta_0, \beta_1, \dots, \beta_p$ neznáme, musíme ich odhadnúť. S ich odhadmi $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ potom vieme pomocou modelu

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p$$

odhadnúť pre konkrétne hodnoty nezávislé premenných x_1, \dots, x_p hodnotu \hat{y} . Tie sa odhadujú použitím metódy najmenších štvorcov (MNS), ktorá spočíva v tom, že sa minimalizuje reziduálny súčet štvorcov, vyjadrený ako

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \dots - \hat{\beta}_p x_{ip})^2.$$

3.2.1. Lineárna regresia s kvalitatívnymi premennými

V predchádzajúcej časti bola vysvetlená lineárna regresia vo všeobecnom tvare, kde sú všetky premenné kvantitatívne. Z praxe je však jasné, že nie vždy to je takýto prípad. Veľmi často sa v dátach zobrazujú veličiny ako pohlavie, krajina pôvodu alebo profesia. Sú to veličiny, ktoré sa nedajú vyjadriť číslom a nazývajú sa kvalitatívne.

Ak máme kvalitatívnu veličinu, ktorá nadobúda iba dve rôzne hodnoty, vytvorí sa pre ne umelá premenná definovaná ako

$$x_i = \begin{cases} 1 & \text{ak má } i\text{-té pozorovanie prvú hodnotu kvalitatívnej veličiny,} \\ 0 & \text{ak má } i\text{-té pozorovanie druhú hodnotu kvalitatívnej veličiny.} \end{cases}$$

Model lineárnej regresie má potom podobu

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{ak } x_i = 1 \\ \beta_0 + \epsilon_i & \text{ak } x_i = 0. \end{cases}$$

V prípade troch hodnôt kvalitatívnej veličiny vyzerá model podobne a to tak, že sa znova vytvorí umelé premenné

$$x_{i1} = \begin{cases} 1 & \text{ak má } i\text{-té pozorovanie prvú hodnotu kvalitatívnej veličiny,} \\ 0 & \text{ak } i\text{-té pozorovanie nemá prvú hodnotu kvalitatívnej veličiny.} \end{cases}$$

$$x_{i2} = \begin{cases} 1 & \text{ak má } i\text{-té pozorovanie druhú hodnotu kvalitatívnej veličiny,} \\ 0 & \text{ak } i\text{-té pozorovanie nemá druhú hodnotu kvalitatívnej veličiny.} \end{cases}$$

Výsledný model tak bude v tvare

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{ak } x_{i1} = 1 \\ \beta_0 + \beta_2 + \epsilon_i & \text{ak } x_{i2} = 1 \\ \beta_0 + \epsilon_i & \text{ak } x_{i1} = 0 \wedge x_{i2} = 0, \end{cases}$$

kde v prvom prípade má i -té pozorovanie prvú hodnotu kvalitatívnej veličiny, v druhom prípade má druhú hodnotu a v poslednom tretiu, tzv. referenčnú hodnotu (*ang. baseline*). Pre kvalitatívne veličiny s viacerými možnými hodnotami, ako napríklad spomínané krajiny pôvodu, sa model formuluje analogicky.

3.2.2. Lineárna regresia kompozičných dát s reálnou vysvetľovanou premennou

Ak si kompozíciu $\mathbf{x} = (x_1, x_2, \dots, x_D)$ vyjadrím v jednom zo súradnicových priestorov, je možné na dáta aplikovať štandardnú viacnásobnú lineárnu regresiu, ktorá bola vysvetlená v prechádzajúcom odstavci. Tento model potom klasicky využíva metódu najmenších štvorcov na odhadnutie premennej. Ak je n vysvetľujúcich premenných, potom má model tvar

$$Y_i = \beta_0^{(l)} + \beta_1^{(l)} z_{i,1}^{(l)} + \beta_2^{(l)} z_{i,2}^{(l)} + \dots + \beta_{D-1}^{(l)} z_{i,D-1}^{(l)} + \epsilon_i,$$

pre $i = 1, \dots, n$ a $l = 1, \dots, D$.

Je dôležité si uvedomiť, že dostávame D rôznych modelov mnohonásobnej regresie, nie iba jeden. Zo vzťahu medzi jednotlivými súradnicovými systémami, ktoré sú si navzajom rotáciami plyní, že $\beta_0^{(l)} = \beta_0$ pre každé $l = 1, \dots, D$.

Pomocou metódy najmenších štvorcov tiež dostaneme koeficient determinácie, vyjadrený ako

$$R^2 = \frac{RSS}{\sum_{i=1}^n (Y_i - \bar{Y})^2},$$

kde \bar{Y} je aritmetický priemer Y_i pre $i = 1, \dots, n$. Rovnako dostaneme aj F -štatistiku

$$F = \frac{R^2}{1 - R^2} \frac{n - D}{D - 1},$$

ktorá má pri platnosti nulovej hypotézy F -rozdelenie s $D - 1$ a $n - D$ stupňami voľnosti. Taktiež to bude platiť o hodnotách regresných koeficientov, ďalších kvalitatívnych a kvantitatívnych (nekompozičných) premenných, ktoré sa v modeli vyskytujú. Ako výsledok dostaneme tabuľku, ktorá obsahuje odhadované parametre $\beta_0^{(l)}, \beta_1^{(l)}, \beta_2^{(l)}, \dots, \beta_{D-1}^{(l)}$ spolu s ostatnými charakteristikami ako sú štandardné odchýlky, hodnoty T -štatistiky a príslušné p -hodnoty. Znova je dôležité vyzdvihnúť, že týchto charakteristík dostávame v rovnakom počte, ako sú odhadované parametre, teda táto tabuľka je vlastne zhrnutím informácie z D modelov.

Kapitola 4

Analýza dát

Po zoznámení so štatistickými metódami, je možné priamo prejsť na analyzovanie dát. V tejto kapitole som čerpala zo zdrojov [3, 6, 2].

4.1. Predstavenie dát

4.1.1. Dáta na analyzovanie štruktúry typu aktivity

Dáta boli poskytnuté jedným cyklistom, ktorý používal aplikáciu Training Peaks na monitorovanie svojej športovej činnosti. V prvom rade bolo potrebné dáta spracovať do podoby, aby bolo možné ich analyzovať. Jednotlivé športovcove aktivity (tréningy a preteky) boli rozdelené do 5 skupín, ktoré boli definované v prvej kapitole, a to Easyride, Endurance, Intervaly, Preteky - kritérium, Preteky - Road race. V pôvodnej myšlienke som chcela zahrnúť aj kategóriu Recovery, ale keďže je to málo výkonnostný tréning, zahrňovali dáta veľa nulových hodnôt vo vyšších silových zónach, ktoré by bolo ťažko logaritmicky transformovať. Pre účely tejto práce nebola táto kategória potrebná, preto som sa rozhodla ju z analýzy vylúčiť. Ďalej využívam model šiestich silových zón. Čas každej z analyzovaných kategórií bol rozdelený podľa toho, koľko času strávil cyklista v každej z nich a bol zapísaný v sekundách. Spolu s každou aktivitou bol taktiež spísaný jej celkový čas, aké mala TSS skóre a koľko kilokalórií cyklista spálil. V prípade pretekov boli zaznamenané aj informácie o tom, či bol pretek úspešne ukončený alebo nie. Výsledné dáta sú uvedené v obrázku 4.1 a v prílohe.

zona1	zona2	zona3	zona4	zona5	zona6	tss	kcal	typ
8122	4146	832	323	532	659	191	3494	pretekRR
1799	1184	488	235	536	357	206	2504	pretekRR
4410	1309	508	282	627	1208	182	1159	pretekRR
4705	1453	625	341	752	1206	171	2107	pretekRR
2289	1561	957	571	830	482	119	1824	pretekRR
5627	1646	642	361	842	1031	154	2367	pretekRR
4295	1888	799	417	915	1446	185	2477	pretekRR
2429	886	552	327	569	791	99	1577	pretekRR
2381	1048	491	261	649	1372	136	1710	pretekRR
3256	1502	729	419	947	1540	167	2237	pretekRR
5663	1800	1363	724	1124	1994	206	3032	pretekRR
1612	428	295	252	174	574	77	773	pretekKRIT
652	136	95	76	79	332	34	407	pretekKRIT
2480	366	355	327	256	1040	77	1156	pretekKRIT
2375	445	301	296	309	806	71	1037	pretekKRIT
1667	311	128	79	244	725	89	765	pretekKRIT
1660	351	136	82	220	874	94	1031	pretekKRIT
2410	581	256	159	376	898	72	1443	pretekKRIT
1009	173	68	44	116	461	33	489	pretekKRIT
1883	627	299	173	402	838	91	1213	pretekKRIT
2042	371	148	87	233	655	58	1023	pretekKRIT
1336	475	212	131	250	503	63	721	pretekKRIT
1568	575	307	209	458	746	93	1036	pretekKRIT
1747	267	106	67	162	586	52	654	pretekKRIT
2536	507	214	134	338	1217	103	1244	pretekKRIT
5399	1366	595	337	839	2191	237	2788	pretekKRIT
1705	362	169	111	257	674	84	612	pretekKRIT
3885	3293	184	34	13	3	58	528	easyride
1967	3034	281	80	89	28	53	996	easyride
1210	3371	556	115	83	60	64	1129	easyride
3620	3901	558	93	68	32	75	1213	easyride
1272	1762	236	172	142	43	44	670	easyride
3182	2193	982	296	273	154	79	1072	easyride
1250	6051	1175	222	248	134	117	1943	easyride
2115	5052	1126	198	166	130	106	1700	easyride
2414	5756	815	120	86	92	120	1915	easyride
3319	3219	2269	694	826	574	163	2481	interval
1673	4003	2944	759	599	211	160	2105	interval
4303	4169	1331	499	515	147	139	2373	interval
1937	7077	1144	154	92	283	152	2293	interval
1802	7153	1339	193	366	223	149	2399	interval
4298	3627	695	723	755	421	151	2075	interval
6783	4523	1529	1251	679	251	173	3423	endurance
5352	4582	2827	919	764	253	160	2741	endurance
7727	4509	3925	1181	752	235	227	2989	endurance
5745	3637	3417	1104	725	267	193	3075	endurance

Obrázek 4.1: Dáta o tréningoch

4.1.2. Dáta na analyzovanie polarizovaného tréningového plánu

V tejto časti sú predstavené dáta, ktorými sa dá lepšie pozrieť na to, či sa oplatí mať štruktúrovaný tréning, v mojom prípade polarizovaný, alebo nie. Dáta sú z rovnakého zdroja, od rovnakého športovca zo softvéru Training Peaks. A to dáta zo 16. novembra 2020 do 10. januára 2021 a od 15. novembra 2021 do 9. januára 2022. Údaje sú rovnaké ako v predchádzajúcich dátach, teda čas strávený v zónach, spálené kálorie a hodnoty TSS. Tu sú však zozbierané po týždni, teda hodnoty sú sčítané z celých týždňov. Sledovanie trvalo dokopy 8 týždňov, pretože je to odporúčaný čas trvania jedného mezocyklu. Analyzujem a porovnávam dve časové obdobia v časovom rozmedzí jedného roka. Obe sa nachádzajú na presne tom istom mieste tréningového roka, aby sa vylúčili ostatné vplyvy (napríklad to, že v strede sezóny je cyklista prirodzene silnejší). Rozdielom týchto dvoch období je to, že v prvom tréningy neboli štruktúrované, zatiaľčo v druhom bola zachovaná systematická štruktúra polarizovaného tréningu. V tomto prípade mám k dispozícii ďalšiu hodnotu. To je hodnota FTP pred začatím cyklu a po ňom. V prvom to bolo 312 W pred a 320 W po a v druhom 325 W pred a 336 W po. Váha cyklistu pritom bola v oboch týchto prípadoch rovnaká, preto sa nemusí prepočítavať na W/kg. Je dôležité si uvedomiť, že pôvodne sa polarizovaný tréning trénuje v 3 zónach. Tieto zóny sa špeciálne prepočítavajú pre cyklistu, aby vedel, aké wattý má na tréningu vyprodukovať. Keďže mojim cieľom je len porovnať dve obdobia, pre tento účel nie je potrebné zóny prepočítavať a je možné pokračovať s našimi už známymi šiestimi zónami.

week	zona1	zona2	zona3	zona4	zona5	zona6	tss	kcal	obdobie
1	3780	9720	2520	540	180	180	180	246	4448 prvý
2	26945	29239	1530	300	128	311	608	608	8419 prvý
3	16425	27801	3144	482	117	173	548	548	9242 prvý
4	26257	34082	5945	1724	367	501	852	852	8092 prvý
5	26100	21660	6180	720	180	240	580	580	5992 prvý
6	19768	24345	4013	794	144	105	555	555	6717 prvý
7	25847	23447	6240	1980	180	1	625	625	10457 prvý
8	13075	14664	5274	2139	556	500	454	454	6045 prvý
1	16800	27060	960	900	960	300	530	530	8089 druhý
2	21480	16800	240	420	960	1020	465	465	7266 druhý
3	14220	19980	480	360	960	540	413	413	6908 druhý
4	10778	14406	190	36	137	227	248	248	4582 druhý
5	29020	18253	1593	501	749	482	445	445	8242 druhý
6	18651	27271	386	472	1198	809	544	544	9078 druhý
7	18409	14358	924	1007	726	170	402	402	6874 druhý
8	12055	9537	870	2642	1241	427	465	465	5716 druhý

Obrázek 4.2: Dáta z dvoch období

4.2. PCA analýza

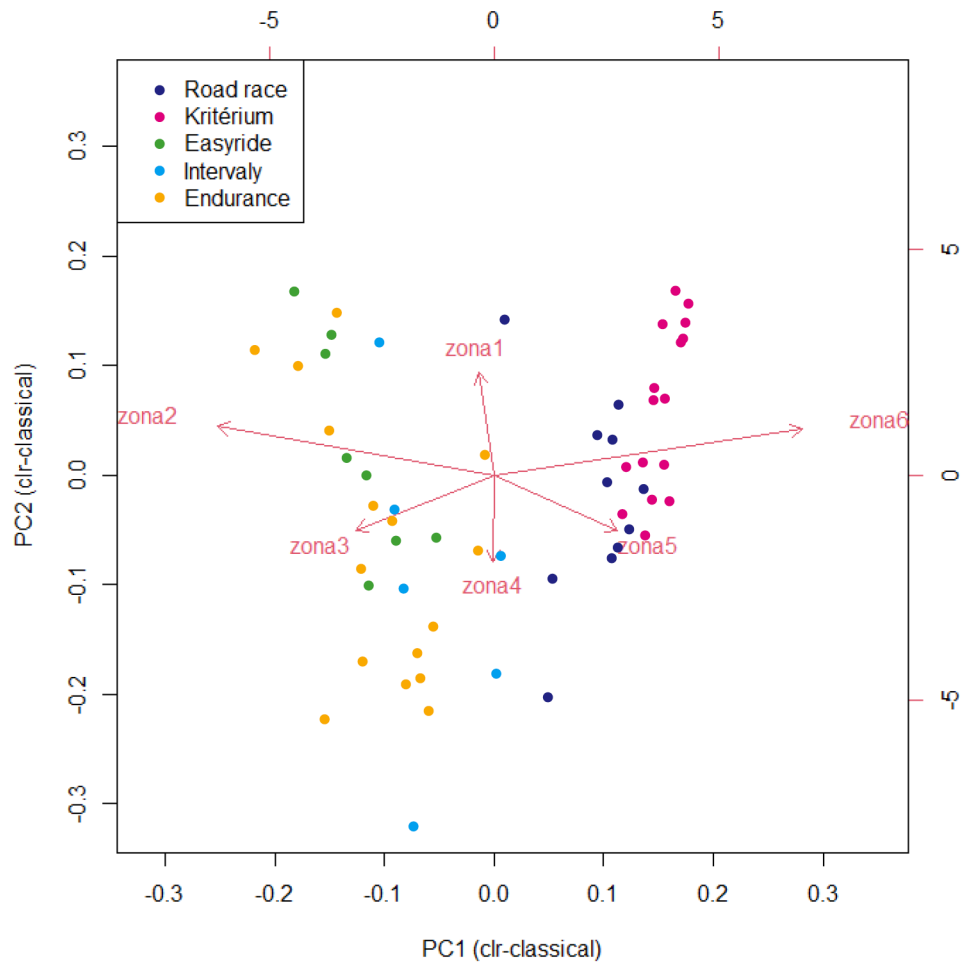
V tejto podkapitole som na dáta použila funkciu `pcaCoDa` z balíka *RobCompositions* v softvéri R. Skontrolovala som, aké percento nášho dátového súboru vysvetľujú hlavné komponenty, v mojom prípade dva. Ďalej som si rôzne označila jednotlivé typy, aby boli ľahko rozoznateľné v grafe, a vykreslila biplot. Cieľom bolo zistiť, či relatívna štruktúra tréningov ozaj odpovedá ich očakávanej obťažnosti (vyjadrenej pomocou jednotlivých silových zón).

```
#PCA analýza
dokopy<-read_excel("dokopy.xlsx", col_names = TRUE)
dokopy<-dokopy[,-c(7,8,9)]
dokopy<-as.data.frame(dokopy)
faz_clas <- pcaCoDa(dokopy, method = "classical")
biplot(faz_clas)
```

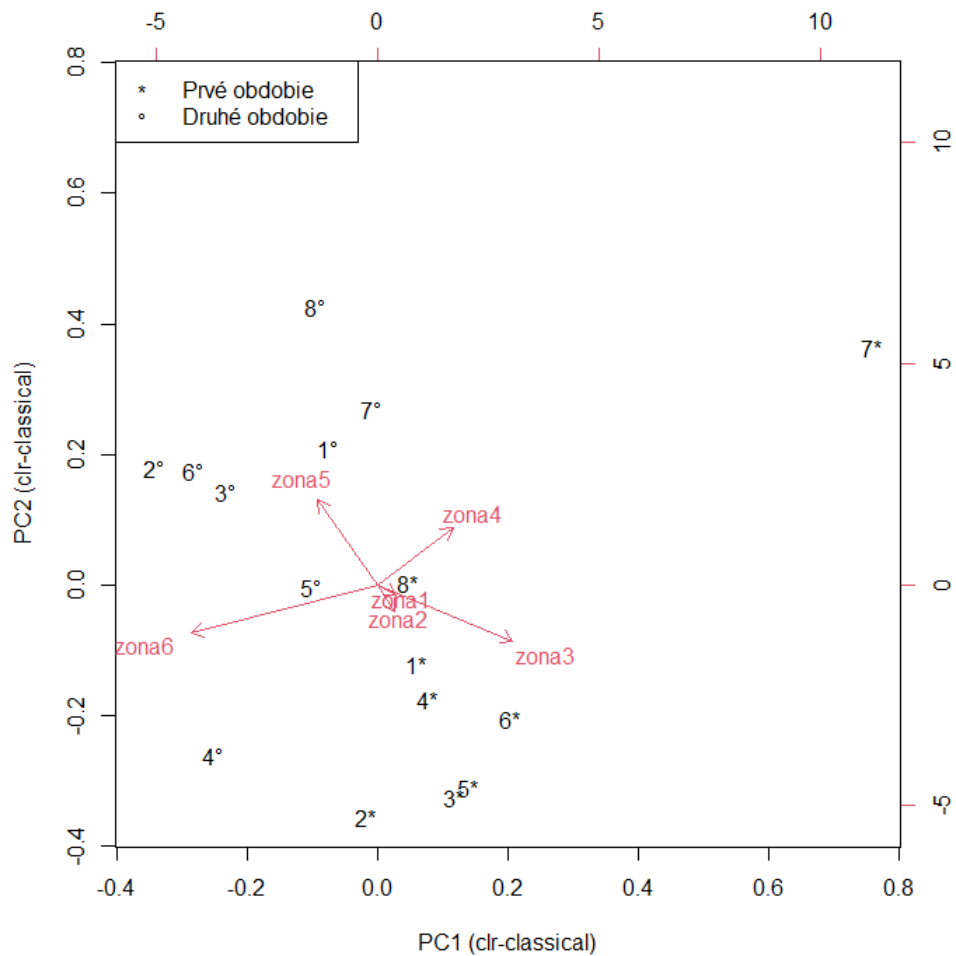
```

#PCA analýza
obdobia<-read_excel("obdobia.xlsx", col_names = TRUE)
as.data.frame(obdobia)
obdobia1<-obdobia[,-c(1,8,9, 10)]
obdobianove<-as.data.frame(obdobia1)
faz_claso <- pcaCoDa(obdobianove, method = "classical")
biplot(faz_claso,
xlabs = c("1*","2*","3*","4*","5*","6*","7*","8*","1°","2°","3°","4°","5°","6°","7°","8°"))
legend(x = "topleft",
legend = c("Prvé_obdobie","Druhé_obdobie"),
lwd = 2, lty = c(0,0),
pch = c("*","°"))

```



Obrázek 4.3: PCA - typy tréningov



Obrázek 4.4: PCA - obdobia

Hlavné komponenty v prvom biplote vysvetľujú 91,3% pôvodného dátového súboru, z toho prvý 80,2% a druhý 11,1%. Je na ňom možné vidieť, že uhol medzi podielom šiestej zóny a prvej je takmer 90 stupňov, teda korelačný koeficient medzi nimi je blízko nuly. To však vzhľadom k singularnej variančnej matici clr transformovaných dát nie je veľmi smerodajné. Skôr sa dá zamerať na ohodnotenie možnej proporcionality medzi dvojicami zložiek, ktorá by tu

mohla byť vyššia medzi dvojicami susediacich zón. Zóna 6 a zóna 2 sú určené dlhšími šípkami, takže ich smerodajná odchýlka je väčšia než u iných zón a majú teda väčší vplyv na celkovú dátovú štruktúru. Z jednotlivých pozorovaní sa vytvorili zhluky. V pravej časti grafu sa zlúčili Road Race a Kritérium. Je možné vidieť vysoký (relatívny) podiel času v šiestej zóne. To sa dalo očakávať už len z faktu, že preteky sú náročné a preto podiel šiestej zóny bude výraznejší než pri tréningoch. Prekvapivý je však rozdiel podielu prvej zóny. Kým v Road race iba jedno z jeho pozorovaní má väčší podiel, pričom ostatné majú priemerný a najmä podpriemerný, u kritérií má prvá zóna väčší význam (samozrejme stále pri dominancii zón 5 a 6). Je pekne vidieť, že v kritériách je to buď maximum alebo minimum. Nakoľko sú kritéria rýchle preteky s priebežným bodovaním, cyklista nemá čas ísť v stredných intenzitách. Na druhej strane Road race sú dlhé preteky, vystupuje tu už aj zóna 3 a 4 a má o čosi menšie zastúpenie zóna 6 než u kritérií. To znamená, že cyklista celý čas produkuje vysoké watt. Tréningy sa skôr zobrazujú v ľavej časti grafu, takže intenzita je tu omnoho menšia. Najväčšie zastúpenie zóny 2 a zóny 1 má Easyride. Pri Endurance dominuje zóna 3 a zóna 4. To priamo vychádza z definícií tréningov. Kým Easyride je tréning na nižšej úrovni, Endurance trénuje rýchlejšie tempo a vytrvalosť. V prípade Intervalov mám k dispozícii iba 8 pozorovaní. Z grafu vidno veľkú variabilitu tohto typu tréningu. Tá môže byť spôsobená buď nedostatkom pozorovaní, alebo faktom, že Intervaly sa dajú trénovať v rôznej frekvencii a rôznych opakovaníach.

Hlavné komponenty v druhom biplote, teda pri pozorovaniach z dvoch období, vysvetľujú 84,3% pôvodných dát, z toho prvý 66,5% a druhý 17,8%. Hneď je vidno, že sa obdobia rozdelili na dve časti. Prvé sa nachádza v spodnej časti grafu a druhé vo vrchnej časti. Je to tým, že v druhom období, mali omnoho väčšie zastúpenie zóna 5 a 6. Preto obe zóny, lebo podľa teórie sa polarizovaný tréning trénuje len v troch, teda piata a šiesta by po približných výpočtoch na trojzónový model, bola jedna zóna. V prvom období naopak vidno veľké zastúpenie zóny 3. Ďalej sa tu nachádzajú dve odľahlé hodnoty a to siedmy týždeň v prvom období a ôsmy v druhom. Vysvetlenie som hľadala v pôvodných dátach. V siedmom týždni sa

cyklista úplne vyhol šiestej zóne, čo spôsobilo takú odľahlosť a v ôsmom týždni naopak nebola zachovaná štruktúra polarizovaného tréningu a cyklista strávil omnoho viac času v štvrtej zóne. To v konečnom dôsledku môže ovplyvniť aj efektivitu polarizovaného tréningu a skresliť výsledky.

4.3. Lineárna regresia

Podľa kapitoly 3.2.2. má z lineárnej regresie vyjsť 6 modelov, podľa šiestich tréningových zón. Keďže pracujem s kompozičnými dátami, v prvom rade ich treba pripraviť tak, aby bol na ne model aplikovateľný. Najprv som z mojich dát vybrala iba stĺpce, ktoré popisovali jednotlivý čas v zónach, teda moje kompozičné dáta, ktoré som následne transformovala funkciou `pivotCoord` pre získanie príslušných izometrických lodpodielových súradníc (v tomto prípade nazývaných tiež pivotové súradnice). Urobila som tak pre každú zónu ako primárnu súradnicu a následne na každú zo šiestich tabuliek aplikovala model lineárnej regresie. Z každého zo šiestich modelov som zobrala informáciu o prvej súradnici a doplnila do jedného modelu spolu s informáciami o kvalitatívnej premennej, type tréningu.

```
#R-kód
#načítanie dát z dvoch období
obdobia<-read_excel("obdobia.xlsx", col_names = TRUE)
as.data.frame(obdobia)
obdobia1<-obdobia[,c(1,8,9,10)]
obdobianove<-as.data.frame(obdobia1)

#pivot transformácia
ob_zona1<-pivotCoord(obdobianove)
ob_zona2<-pivotCoord(obdobianove, pivotvar = 2)
ob_zona3<-pivotCoord(obdobianove, pivotvar = 3)
ob_zona4<-pivotCoord(obdobianove, pivotvar = 4)
ob_zona5<-pivotCoord(obdobianove, pivotvar = 5)
ob_zona6<-pivotCoord(obdobianove, pivotvar = 6)
```

Dostala som 6 pivotových súradnicových systémov, ktoré som spojila s reálnymi premennými, t.j. stresovým skóre a spálenými kalóriami

```

colnames(ob_zona1)<- (c("p1", "p2", "p3","p4","p5"))
zona1obd<-cbind(ob_zona1, obdobia[c(8,9,10)])
colnames(ob_zona2)<- (c("p1", "p2", "p3","p4","p5"))
zona2obd<-cbind(ob_zona2, obdobia[c(8,9,10)])
colnames(ob_zona3)<- (c("p1", "p2", "p3","p4","p5"))
zona3obd<-cbind(ob_zona3, obdobia[c(8,9,10)])
colnames(ob_zona4)<- (c("p1", "p2", "p3","p4","p5"))
zona4obd<-cbind(ob_zona4, obdobia[c(8,9,10)])
colnames(ob_zona5)<- (c("p1", "p2", "p3","p4","p5"))
zona5obd<-cbind(ob_zona5, obdobia[c(8,9,10)])
colnames(ob_zona6)<- (c("p1", "p2", "p3","p4","p5"))
zona6obd<-cbind(ob_zona6, obdobia[c(8,9,10)])

```

Ďalším postupom bolo na každú zo šiestich tabuliek aplikovať model lineárnej regresie. V každom z modelov sa dalo pozorovať rovnakú smerodajnú odchýlku, koeficient determinancie či F -štatistiku. Následne som zo všetkých zobrala informáciu o prvej súradnici a doplnila do jedného modelu spolu s informáciami o kvalitatívnej premennej, type tréningu, ktoré som zapísala do jednej tabuľky manuálne. Rovnaký postup som použila v prípade dát, kde porovnáваме typy tréningov. Obdobný R-kód sa nachádza v prílohe.

```

#šesť modelov lineárnej regresie pre tss
lm1obd<-lm( tss ~ p1 + p2 + p3 + p4 + p5 + obdobie , data = zona1obd)
summary(lm1obd)
lm2obd<-lm( tss ~ p1 + p2 + p3 + p4 + p5 + obdobie , data = zona2obd)
summary(lm2obd)
lm3obd<-lm( tss ~ p1 + p2 + p3 + p4 + p5 + obdobie , data = zona3obd)
summary(lm3obd)
lm4obd<-lm( tss ~ p1 + p2 + p3 + p4 + p5 + obdobie , data = zona4obd)
summary(lm4obd)
lm5obd<-lm( tss ~ p1 + p2 + p3 + p4 + p5 + obdobie , data = zona5obd)
summary(lm5obd)
lm6obd<-lm( tss ~ p1 + p2 + p3 + p4 + p5 + obdobie , data = zona6obd)
summary(lm6obd)

#šesť modelov lineárnej regresie pre kcal
lm1obk<-lm( kcal ~ p1 + p2 + p3 + p4 + p5 + obdobie , data = zona1obd)
summary(lm1obk)
lm2obk<-lm( kcal ~ p1 + p2 + p3 + p4 + p5 + obdobie , data = zona2obd)
summary(lm2obk)
lm3obk<-lm( kcal ~ p1 + p2 + p3 + p4 + p5 + obdobie , data = zona3obd)
summary(lm3obk)
lm4obk<-lm( kcal ~ p1 + p2 + p3 + p4 + p5 + obdobie , data = zona4obd)
summary(lm4obk)
lm5obk<-lm( kcal ~ p1 + p2 + p3 + p4 + p5 + obdobie , data = zona5obd)
summary(lm5obk)
lm6obk<-lm( kcal ~ p1 + p2 + p3 + p4 + p5 + obdobie , data = zona6obd)
summary(lm6obk)

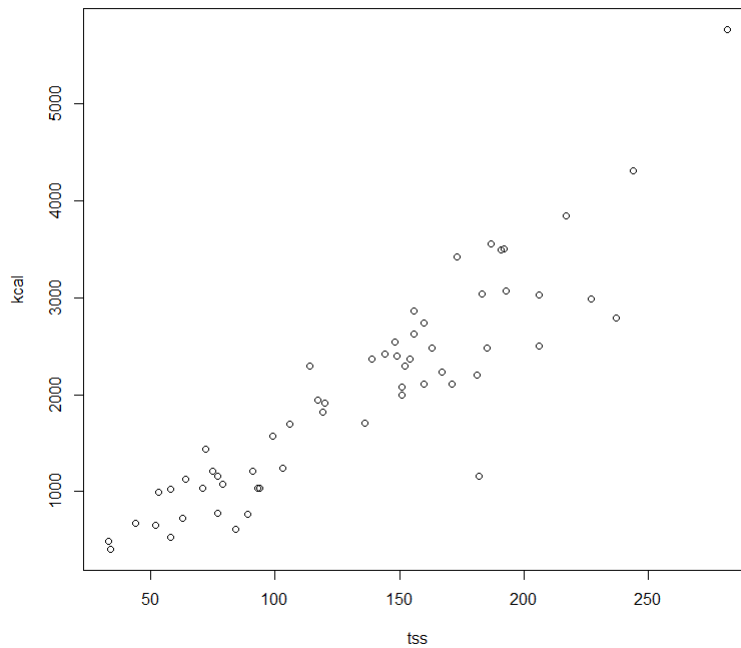
```

...1	Estimate	Std. Error	t value	Pr(> t)	...1	Estimate	Std. Error	t value	Pr(> t)
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
Intercept	53.0	33.6	1.58	0.122	Intercept	440.	586.	0.751	0.456
zona1	6.13	12.4	0.491	0.626	zona1	-21.2	217.	-0.097	0.923
zona2	-6.65	13.7	-0.485	0.630	zona2	277.	239.	1.16	0.251
zona3	26.0	22.3	1.17	0.249	zona3	45.3	388.	0.117	0.908
zona4	-39.3	22.5	-1.75	0.0868	zona4	-377.	392.	-0.961	0.341
zona5	12.9	21.7	0.595	0.555	zona5	20.7	377.	0.055	0.957
zona6	0.940	13.2	0.071	0.944	zona6	54.5	230.	0.236	0.814
endurance	106.	17.1	6.18	0.000000123	endurance	2008.	297.	6.75	0.0000000161
intervaly	74.3	22.9	3.24	0.00213	intervaly	1246.	399.	3.12	0.00301
kriterium	3.86	40.0	0.097	0.923	kriterium	245.	696.	0.352	0.726
roadrace	83.9	32.0	2.62	0.0118	roadrace	1380.	558.	2.47	0.0169

Obrázek 4.5: Výsledné modely typov tréningov pre vysvetľované premenné (TSS - vľavo, Kcal - vpravo)

Pri sledovaní p-hodnôt je možné vidieť významnú hodnotu pri štvrtej zóne v prípade stresového skóra. Odtiaľ vidieť, že zvýšenie dominancie štvrtej zóny povedie k zníženiu stresového skóra. To je prekvapivé, pretože štvrtá zóna je zóna vyššej intenzity, teda by prirodzene malo stresové skóre rásť. Čo sa však dá povedať je to, že tvorba TSS exponenciálne rastie smerom k šiestej zóne. Teda ak zväčšíme podiel v štvrtej, zmenší sa podiel v šiestej a piatej, teda sa môže vyprodukovať menej TSS. Čo sa týka ostatných zón, podiel času v nich strávený už nie je tak štaticky významný. Dôležité sa však ukazujú byť typy tréningu. Pri Endurance vyšla p-hodnota veľmi malá. Model ukazuje na to, že pri tréningu Endurance cyklista vyprodukuje 106 TSS viac ako pri Easyride (referenčná premenná, ktorá je poísaná parametrom pri absolútnom člene), ak ostatné premenné ostanú nezmenené. V prípade Intervalov je to o 74 TSS viac a v prípade Road Race je to o 84 viac. Hodnoty vychádzajú podľa očakávania, pretože Easyride, ku ktorému sú tieto typy porovnávané, je tréning najnižšej intenzity. Čo je však nové je to, že množstvo TSS pri intervaloch je podobné ako pri dlhých pretekoch a oboje je nižšie než pri Endurance. V prípade spálených kalórií v roli závislej premennej nevyzerá byť podiel v jednotlivých zónach taký významný, ako znova to, o aký typ tréningu ide. Pri Endurance vyšlo, že cyklista spáli o 2008 viac kalórií ako pri ľahkom tréningu. Pri intervaloch a dlhých pretekoch sa znova ukazuje podobné navýšenie voči ľahkému tréningu. Čo sa týka kritérii, v oboch prípadoch tu vyšla veľká p-hodnota, ako jedinému z typov tréningov. Je teda možné vyvodiť, že pri type tréningu resp. preteku nejde len o to, koľko času v ktorej zóne cyklista

strávi. Napríklad pri preteku môže mať veľký vplyv na TSS aj kalórie dodatočný stres pridaný situáciou. Sledujúc obe tabuľky súčasne, je vidieť podobnosti. Preto som vyrátala koreláciu premenných TSS a spálených kcal.



Obrázek 4.6: Korelácia TSS a Kcal

Ako už nasvedčoval model lineárnej regresie, medzi TSS a kcal vyšiel vysoký korelačný koeficient a to 0,91. Teda čím sa telo viac unaví a vyprodukuje viac stresového skóre, tým viac kalórií spáli.

...1	Estimate	Std. Error	t value	Pr(> t)	...1	Estimate	Std. Error	t value	Pr(> t)
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
Intercept	107.	278.	0.47	0.649	Intercept	1790.	2725.	0.657	0.528
zona1	130.	121.	1.08	0.31	zona1	1012.	1447.	0.699	0.502
zona2	2.24	140.	0.016	0.988	zona2	-721.	1676.	0.43	0.677
zona3	-157.	92.4	1.70	0.124	zona3	-1624.	1106.	-1.47	0.176
zona4	191.	119.	1.61	0.143	zona4	939.	1425.	0.659	0.526
zona5	-183.	162.	-1.13	0.288	zona5	-454.	1943.	-0.233	0.821
zona6	17.0	37.5	-0.455	0.66	zona6	-593.	448.	-1.32	0.218
prvý	92.1	259.	0.356	0.73	prvý	1351.	3099.	0.436	0.673

Obrázek 4.7: Výsledné modely dvoch období pre vysvetľovanú premennú (TSS - vľavo, Kcal - vpravo)

V ďalšom kroku som previedla rovnakú analýzu pre dáta rozdelené podľa štruktúry tréningového plánu - prvé obdobie tu odpovedalo neštruktúrovanému plánu a druhé štruktúrovanému. Cieľom bolo opäť zistiť ich vzťah k TSS, resp. spáleným kalóriám.

V tomto prípade vyšli celkom veľké p-hodnoty. Najmenšia hodnota vyšla pri zóne 3 v oboch prípadoch, pri vzniknutom stresovom skóre aj pri spálených kalóriách. Obe tabuľky nasvedčujú tomu, že ak sa zväčší podiel času stráveného za týždeň v tretej zóne, zmenší sa počet TSS a spálených kalórií. Avšak rozdiel medzi prvým a druhým obdobím sa tu nedá pozorovať. Odpovedajúca p-hodnota pri kategoriálnej premennej je príliš veľká, aby bola štatisticky významná.

4.4. Absolútny rozdiel v porovnávaní dvoch období

Ako som spomínala v odstavci o dátach, cyklistom boli poskytnuté údaje o hodnote FTP na začiatku prvého aj druhého obdobia a aj na konci. V oboch prípadoch sa hodnota FTP zvýšila, čo dáva zmysel, pretože v tom období je tzv. zimná príprava, ktorej cieľom je zosilniť telo a nastaviť na sezónu. Absolútne rozdiely sú však veľmi rozdielne. Kým v prvom období je to $320 - 312 = 8$ wattov, v druhom je to $336 - 325 = 11$ wattov. Môže sa zdať, že 3 watty nie sú veľký rozdiel, ale v cyklistike sú veľmi významné. Tiež môže byť zaražujúce, že na začiatku obdobia bolo raz 312 wattov a po roku až 325, keď je to ten istý čas v roku. To však nie je prekvapujúce, pretože ak cyklista konzistentne trénuje, rok čo rok sa jeho FTP zvyšuje a jeho forma a sila sa zväčšujú. Tento rozdiel v našom prípade nie je podstatný, pretože skúmame len absolútny prírastok wattov v oboch prípadoch.

Čo však možno stojí za zmienku je fakt, že čím väčšie FTP cyklista má, tým ide ťažšie hore. To znamená, že je jednoduchšie zvýšiť svoje FTP zo 120 wattov na 160 wattov, než z 320 wattov na 360 wattov. Je to preto, že telo sa dostáva bližšie k svojej maximálnej sile, ktorú je biologicky schopné vyprodukovať. Ak to teda zohľadníme pri našej analýze, dáva to absolútnemu rozdielu 11 wattov ešte väčšiu váhu. V konečnom dôsledku sú tie 3 watty ešte dôležitejšie, než by sa na prvý pohľad mohlo zdať.

Napriek tomu, ako sa ukázalo v predchádzajúcej kapitole, na TSS a počtu spálených kalórií sa z hľadiska relatívnej štruktúry tréningu žiaden signifikantný rozdiel medzi obidvoma obdobiami neprejavil. Na druhú stranu, PCA poukázala na ich očividnú kvalitatívnu odlišnosť čo sa týka distribúcie tréningových zón.

Záver

Analýza hlavných komponentov priniesla pekný pohľad do štruktúry jednotlivých typov tréningov. Bolo zaujímavé vidieť rôzne hodnoty, ktoré sa priamo dali vysvetliť poznatkami o tréningoch. Lineárna regresia ukázala, ako veľmi záleží na type tréningu pri počte vyprodukovaného stresového skóra a spálených kalórii. Nezáleží teda až tak na tom, aký podiel má ktorá zóna, ale na tom, na čo sa tréning zameriava. Pri analýze hlavných komponentov sa dalo sledovať priame súvislosti a podobnosti medzi jednotlivými typmi tréningov už z tak malej vzorky, akú som mala poskytnutú. To isté sa dá povedať o lineárnej regresii. Určite by boli výsledky presnejšie s väčším vzorkom pozorovaní a od viacerých cyklistov, no i v tomto prípade už boli viditeľné pekné štruktúry.

Prechádzajúc na skúmanie dvoch období znova vychádzali zaujímavé výsledky. Analýza hlavných komponentov krásne rozdelila obdobia a priamo ukázala, že v druhom bol podiel šiestej a piatej zóny výrazne vyšší a tréning bol intenzívnejší. To je vlastne myšlienka polarizovaného tréningu. Na druhú stranu lineárna regresia pri porovnávaní dvoch období a skúmaní, ktoré je lepšie, veľmi nepomohla, čo mohlo, byť vedľa pomerne malého rozsahu dátového súboru, následkom možného výskytu odľahlých hodnôt. V tomto prípade boli však v rámci dát k dispozícii aj hodnoty FTP cyklistu pred a po oboch sledovaných obdobiach. Tým sa ukázalo, že je priamo vidieť efektívnosť polarizovaného tréningu a že sa oplatí trénovať v štruktúre a systematicky. To znamená, že aj keď sa z lineárnej regresie nedal vyvodíť zmysluplný záver, polarizovaný tréning predsa len vyzerá byť efektívny a je dobrým adeptom na ďalšiu analýzu.

Pri písaní tejto práce ma najviac bavilo spoznávať vplyv matematiky na výkon

a tréningy, bavilo ma spisovať dáta a viac sa zoznamovať so športom z iného uhla pohľadu. Jednou z najlepších častí bolo sledovať cyklistu v čase plnenia polarizovaného tréningu a zároveň pozorovať jeho pokroky v priamom prenose a zbierať dáta na následnú analýzu. Výzvou však bolo vybranie správnych štatistických metód, aby sa na tieto dáta hodili a priniesli výsledky. To sa nakoniec podarilo a bola som s tým spokojná. Podľa môjho názoru by čas cyklistu stráveného v jednotlivých silových zónach bol dobrou voľbou na ďalšie detailnejšie a hlbšie analýzy, keďže tréningovanie ide stále dopredu. Spolu so sledovaním cyklistu a reakciami jeho tela na zmeny času v zónach by mohli byť dostupné nové pridané dáta - veličiny na analyzovanie a mohli by byť veľkým prínosom do oblasti športu a tréningovania.

Literatura

- [1] ALLEN, H., COGGAN, A. R., MCGREGOR, S.: *Training and racing with a power meter*. Boulder, Colorado, 2019. ISBN 978-1-937715-93-9
- [2] DUMUID, D., STANFORD, TE., MARTIN-FERNÁNDEZ, J., PEDISIC, Z., MAHER, C., LEWIS, L., HRON, K., KATZMARZYK, P., CHAPUT, J.-P., FOGELHOLM, M., HU, G., LAMBERT, E., MAIA, J., SARMIENTO, O., STANDAGE, M., BARREIRA, T., BROYLES, S., TUDOR-LOCKE, C., TREMBLAY, M. & OLDS, T.: *Compositional data analysis for physical activity, sedentary time and sleep research*. *Statistical Methods in Medical Research*, 27 (12), 3726-3738, 2018.
- [3] FILZMOSER, P., HRON, K., TEMPL, M.: *Applied compositional data analysis*. Springer, Cham, 2018. ISBN 978-3-319-96420-1
- [4] FRIEL, J.: *Tréninková bible pro cyklisty*. Mladá fronta, Praha. 2009. ISBN 978-80-204-2640-6
- [5] GARETH, J., WITTEN, D., HASTIE, T., TIBSHIRANI, R.: *An introduction to statistical learning with applications in R*. Springer, New York, 2021. ISBN 1071614177
- [6] HRON, K., COENDERS, G., FILZMOSER, P., PALAREA-ALBALADEJO, J., FAMĚRA, M., GRYGAR, T. M.: *Analysing pairwise logratios revisited*. *Mathematical Geosciences*, 53, 1643–1666, 2021.