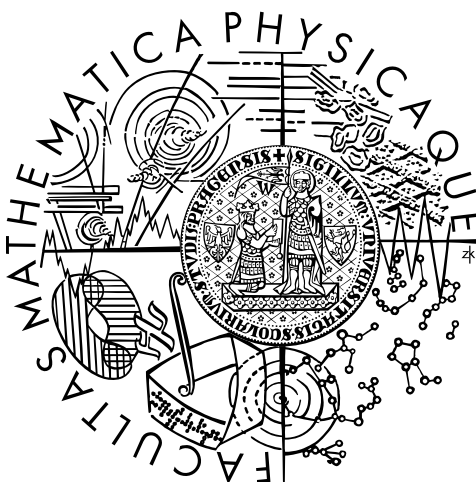


Univerzita Karlova v Praze  
Matematicko-fyzikální fakulta  
**BAKALÁŘSKÁ PRÁCE**



Lenka Kovářová

**Prostředí pro analýzu veteránského rankingu OB**

Vedoucí bakalářské práce:

RNDr. Libor Forst

Středisko infromatické sítě a laboratoří

Studijní program:

Obecná informatika

Informatika

2009

Děkuji vedoucímu bakalářské práce RNDr. Liborovi Forstovi za cenné rady, připomínky a vedení mé práce. Děkuji konzultantovi Mgr. Pavlovi Schlesigerovi za návrh statistických metod. Děkuji rodičům za podporu při studiu.

Prohlašuji, že jsem svou bakalářskou práci napsala samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 5. 8. 2009

Lenka Kovářová

# Obsah

<b>1 Úvod</b> .....	<b>5</b>
1.1 Úvod do problematiky orientačního běhu .....	5
1.2 Orientační běh ještě jednou a trochu jinak .....	6
1.3 Cíl této práce.....	7
<b>2 Ranking</b> .....	<b>8</b>
2.1 Co je to ranking .....	8
2.2 Jak spočítat ranking .....	9
<b>3 Data - výsledky závodů</b> .....	<b>10</b>
3.1 Zdroj dat .....	10
3.2 Formát výsledků závodů.....	11
3.3 Zpracování výsledků závodů.....	12
<b>4 Struktura databáze</b> .....	<b>13</b>
4.1 Závodníci a výsledky závodů .....	13
4.2 Ranking a body za závod.....	14
4.3 Vzorečky v databázi .....	15
<b>5 Uživatelské rozhraní</b> .....	<b>16</b>
5.1 Stránky zobrazující ranking.....	16
<b>6 Statistika</b> .....	<b>17</b>
6.1 Základní pojmy.....	17
6.2 Korelační koeficient .....	18
6.3 Výběrový korelační koeficient .....	18
6.4 Spearmanův korelační koeficient .....	19
6.5 Modifikovaný Spearmanův korelační koeficient .....	19
<b>7 Vzorečky</b> .....	<b>20</b>
7.1 Vzoreček pro výpočet rankingu.....	20
7.2 Vzoreček pro výpočet bodů přidělených za závod.....	21
7.3 Parametrizace vzorečku .....	22
7.4 Motivace pro různé vzorečky .....	22
7.5 Současný český ranking .....	24
7.6 Některé další rankingy.....	25
<b>8 Vyhodnocení</b> .....	<b>26</b>
8.1 Hodnocená data .....	26
8.2 Popis závodů a závodníků .....	26
8.3 Modifikovaný Spearmanův korelační koeficient .....	29
8.4 Podrobnější popis vybraných vzorečků.....	38
<b>9 Závěr</b> .....	<b>42</b>
<b>Literatura</b> .....	<b>43</b>

Název práce: Prostředí pro analýzu veteránského rankingu OB

Autor: Lenka Kovářová

Katedra: Středisko informatické sítě a laboratoří

Vedoucí bakalářské práce: RNDr. Libor Forst

e-mail vedoucího: forst@sisal.mff.cuni.cz

Abstrakt:

Byl proveden důkladný rozbor potenciálních schémat vzorců pro výpočet rankingu veteránských závodníků orientačního běhu a byla navržena jejich možná parametrizace.

Úkolem práce bylo vytvořit prostředí pro výpočet veteránského rankingu, které umožňuje počítat ranking závodníků podle variabilního vzorce podporujícího navržené parametrizace.

Byly shromážděny výsledky závodů orientačního běhu za poslední 4 roky, aby byl zajištěn dostatek dat pro výpočet rankingu. Pro zobrazování vypočteného rankingu byly vytvořeny internetové stránky.

Podle zvolených vzorečků byl spočítán veteránský ranking závodníků orientačního běhu. Pro ověření vhodnosti vzorečků byl zvolen Spearmanův korelační koeficient, který byl dále upraven, aby se lépe hodil k vyhodnocení jednotlivých variant vzorečků a jejich parametrizace.

Klíčová slova:

veteránský ranking, orientační běh, vzorec pro výpočet rankingu, PHP a MySQL, Spearmanův korelační koeficient

Title: Environment for Veteran Orienteering Ranking Analysis

Author: Lenka Kovářová

Department: Network and Labs Management Center

Supervisor: RNDr. Libor Forst

Supervisor's e-mail address: forst@sisal.mff.cuni.cz

Abstract:

I accomplished analyse of potential formula schemes for veteran competitors ranking evaluation and I suggested their parameterisation. Next, I created an environment for the evaluation of veteran ranking that enables the competitors ranking calculation in accordance with the variables formulae that supports the proposed parameterisation and I, therefore, collected the results of orienteering races that were held in the last 4 years to get some data for the ranking evaluation. To display the calculated ranking I constructed a website.

I computed the ranking of veteran orienteering competitors by use of selected formulae and then I suggested a modification of the Spearman correlation coefficient to prove the accuracy of each formula.

Key words:

Veteran ranking, orienteering, formulae for ranking evaluation, PHP and MySQL, Spearman correlation coefficient

# Kapitola 1

## Úvod

### 1.1 Úvod do problematiky orientačního běhu

Již naši předkové v dávných dobách pobíhali po pralese a potřebovali se orientovat. Samozřejmě, že to ještě nebyl správný orientáček, neboli orientační běh, protože ten se, alespoň podle mne, neběhá z povinnosti, ale především pro zábavu. Dnes si přece už nemusíme jít k snídani ulovit opici. Prostě si jen jdeme zaběhat a odreagovat se a v ideálním případě vyhrát.

Takový úvod je prozatím více pro zasvěcence, orientáčky, proto jistě čtenáři méně seznámení s orientačním během uvítají podrobnější vysvětlení o co to vlastně v tom orientačním běhu běží. Tedy jedná se o moderní sportovní disciplínu vytrvalostního charakteru, která má své počátky v roce 1897, kdy byl v Oslu uspořádán první závod. Zpočátku se jednalo jen o místní událost pro pár nadšenců. Mezitím se však tento sport velmi rozšířil a dnes je celosvětově registrováno v International Orienteering Federation (IOF) celkem 60 států a přibližně milion aktivních běžců, přičemž základní disciplíny orientačního běhu se dle IOF dělí na pěší (OB), lyžařský (LOB), s horskými koly (MTBO) a Trail (zaměřuje se na správné čtení mapy, tím umožňuje přístup vozíčkářům).

Jen v České republice je v roce 2009 registrovaných 7096 orientačních běžců. Jejich oblastní rozdělení je však velmi nerovnoměrné. Zatímco ve Středočeské oblasti, včetně Prahy, kde studuji, je přibližně 1500 běžců orientačního běhu, tak v Jihočeské oblasti, včetně Chlumu u Třeboně, kde bydlím, je jen 184 závodníků, aktivních je ještě méně, třeba v našem oddíle je jen 8 aktivních běžců. To ovšem již zabíhám do celkem nepodstatných detailů.

Spíše by Vás mohlo zajímat, jak probíhá takový závod v orientačním běhu. Cílem je nalézt všechny kontrolní body v předepsaném pořadí v co nejkratším čase. Přednostní orientačního běhu je jeho nízká náročnost, jak na výbavu účastníků, stačí Vám buzola, jednoduchý elektronický čip a běžecká obuv, tak také z hlediska pořadatelů, kteří zajistí terén, zázemí, natisknou mapy, připraví kontroly a vyhodnotí závod. Můžeme říct, že se jedná o "zelený sport", pro jeho zanedbatelný vliv na životní prostředí. Je přístupný všem ve věku od 3-99 let (malé děti běhají s doprovodem rodičů). Možná Vás překvapí horní limit věku, je to nadsázka, ale faktem je, že pro mnohé závodníky se stane orientační běh celoživotním koníčkem.

*Pořadatelé závodu* postaví ve vybrané zmapované lokalitě různé tratě, které délkou, počtem kontrol a obtížností odpovídají jednotlivým věkovým kategoriím, přičemž start i cíl je pro všechny kategorie společný. Zajímavostí je, že tratě nejsou vyznačeny po celé své délce, jako u mnohých jiných sportů, ale že tratě jsou souvisle zakresleny jen v mapě a ve skutečnosti se skládají pouze z kontrolních bodů, kterými musí závodník proběhnout.

*Kontroly* se skládají z elektronického zařízení umístěného v malé krabičce na stojanu, který je označen lampionem s charakteristickými bílými a oranžovými nebo červenými trojúhelníky.

*Závodník* je vybaven již zmiňovaným čipem a ten při nalezení své kontroly krátce vloží do elektronického zařízení, čímž uloží do obou čas své přítomnosti v daném bodě. Čip pro závodníka ukládá časy jeho průběhu jednotlivými kontrolními body a zároveň jeho celkový čas.

Pořadatelé závodu dále přidělí každému účastníkovi čas jeho startu. Závodníci zpravidla startují po skupinkách, kde má každý různou trať, v jednodominutových intervalech, s tím, že závodníci stejné kategorie nebývají v bezprostředně po sobě následujících skupinkách. Předchozí opatření nejsou ani tak proto, aby připravily perné chvílky pořadatelům při sestavování rozpisu startujících, ale mají ztlumit shlukování závodníků ze stejných kategorií a zajistit, že jejich výsledné výkony budou dosaženy vlastním úsilím.

*Volba postupu* od jedné kontroly k následující je ponechána na závodníkovi a právě volný výběr cesty je podle mne spolu s celkovou pohodovou atmosférou jedním z hlavních kouzel orientačního běhu. Jste v tom sami, alespoň by jste měli být, a záleží jen na Vás jak si s jednotlivými nástrahami terénu i pořadatelů poradíte. Velkým pomocníkem každého orientačního běžce je *buzola*, která mu rychlým a snadným určením severu pomáhá zorientovat mapu a udržovat směr postupu.

*Běhá se* v nejčastěji v lesích a to je příjemné hned z několika důvodů - pohybujete se v čisté přírodě, můžete si sami zvolit tempo běhu či chůze jak Vám to vyhovuje a poznáte nové krajiny. Navíc běh po měkké půdě šetří klouby, na druhou stranu, zase ne jeden orientační běžec poznal na vlastní kůži, jak snadno si lze na nerovném terénu vykloubit kotník. Všem starostem ovšem dá zapomenout radost z vítězství nebo dobrý pocit z podařeného závodu.

Na první pohled se Vám teď může celá záležitost s orientačním během zdát trochu komplikovaná, ale věřte, že ve skutečnosti je po vyzkoušení si několika závodů vše velmi jednoduché a možná si tento sport také oblíbíte.

## 1.2 Orientační běh ještě jednou a trochu jinak

Orientační běh je individuální sport, kde se závodníci snaží v co nejkratším čase oběhnout všechny pořadatelé zvolené, v lese a na mapě vyznačené kontroly.

Závod orientačního běhu je rozdělen do mnoha různých kategorií. Jednotlivé kategorie jsou odstupňovány podle věku závodníků, občas také podle obtížnosti tratě závodu. Kategorie H21, D21 značí mužské a ženské hlavní kategorie, do těchto kategorií se mohou přihlašovat všichni závodníci bez věkového omezení. Kategorie označené nižším číslem značí juniorské kategorie, ale o ty se při výpočtu veteránského rankingu není třeba příliš zajímat. Kategorie označené vyšším číslem značí veteránské kategorie, číslo udává minimální věk závodníka. Tyto kategorie bývají odstupňované po 5 letech a začínají kategorií pro závodníky od 35 let.

Závodníci se mohou hlásit do různých kategorií závodu tak, aby neporušili věková omezení. Pro veteránské závodníky to znamená, že mohou přihlásit i do kategorie určené pro mladší veteránské kategorie nebo do hlavní kategorie, juniorští závodníci se naopak mohou hlásit do kategorií určených starším juniorům a do hlavní kategorie. V jedné kategorii závodu tedy spolu mohou soutěžit závodníci patřící do různých věkových skupin, v hlavních kategoriích jsou dokonce mohou soutěžit veteránští závodníci spolu s juniorskými.

Závody orientačního běhu mají několik variant startu. Většina závodů má intervalový start. Při intervalovém startu závodníci patřící do stejné kategorie startují v obvykle dvouminutových intervalech. Pořadí ve kterém závodníci dobíhají do cíle není podstatné, výsledné pořadí se určí podle času závodníků.

Handicapový start je další varianta, která bývá používána pro poslední etapu víceetapových závodů. Startovní čas závodníka odpovídá součtu časů, za které běžel předchozí etapy. Při tomto typu startu závodníci dobíhají do cíle v pořadí, které odpovídá jejich celkovému umístění za všechny etapy závodu. Závodníci, kteří nedokončili některou z předchozích etap nebo mají příliš velkou ztrátu na první závodníky, většinou startují intervalovým startem na začátku závodu. Poslední méně častou variantou je hromadný start, kdy startují všichni závodníci stejné kategorie najednou.

Závodníci orientačního běhu jsou lidé soutěživí. Během závodu se snaží předbíhat jeden druhého, i když startovali v rozdílných časech a na místo kde se potkali doběhli každý v jiném čase. Soutěží spolu i po doběhnutí do cíle, kde porovnávají své výkony v lese a s velkým nadšením diskutují dlouhé hodiny nad možnými i nemožnými postupy mezi kontrolami ve snaze najít ten nejlepší. Porovnávají se kdo kdy kde koho předběhl a vymýšlejí kde kdy a koho předběhnou příště.

Ale ani to jim nestačí, a proto vzniknul ranking. Ranking je ohodnocení závodníků, podle kterého je možné sestavit pořadí jednotlivých závodníků. V současné době existuje v České republice jen ranking hlavních kategorií závodů. Ranking závodníka je vypočten jako součet deseti nejvyšších bodů získaných za závody, které se konaly během posledního roku. Tyto body jsou spočítány pomocí podílu času závodníka a průměrného času tří nejrychlejších závodníků, který je vynásobený průměrným rankingem pěti závodníků s nejvyšším rankingem pro zohlednění rozdílné kvality závodů. Na začátku každého měsíce je ranking přepočítán, tím vzniknou nové hodnoty rankingů, které se použijí jak pro průměr rankingů nejlepších závodníků v kategorii při výpočtu bodů přidělených za závody, tak k sestavení aktualizovaného pořadí závodníků. To aby závodníci měli zase co porovnávat, o kolik bodů zlepšili a o kolik míst se posunuli nahoru, no nebyla by to bez toho nuda?

Problém je, že tento ranking je počítán pouze z výsledků hlavních kategorií závodů. Pokud se chtějí veteránští závodníci také účastnit rankingů, jsou nuceni běhat právě v hlavních kategoriích závodů spolu s výrazně mladšími závodníky. V rankingů jsou s nimi také hodnoceni dohromady, a to není vůči veteránským závodníkům úplně fér. A proto by bylo dobré kdyby existoval také veteránský ranking orientačních běžců.

### 1.3 Cíl této práce

Cílem této práce je vytvořit veteránský ranking pro závodníky orientačního běhu a navrhnout vzorečky, podle kterých by se tento ranking mohl počítat. Pro ověření vhodnosti navrhovaných vzorečků je dobré podle těchto vzorečků ranking opravdu spočítat. Aby bylo možné počítat podle téměř libovolně zvoleného vzorečku, je třeba brát na tento požadavek ohled již při návrhu programu a vybudovat mechanismus, který bude dostatečně parametrizovatelný a flexibilní. Je třeba vymyslet celou škálu možných požadavků různých vzorečků pro výpočet rankingů a bodů přidělených závodníkům za závody a umožnit jejich realizaci. Je vhodné, aby systém uměl počítat zároveň podle různých vzorečků. Samozřejmě je třeba navrhnout nějakou metodu pro vyhodnocení korektnosti rankingů spočítaných pomocí jednotlivých vzorečků a na závěr provést vyhodnocení získaných dat.

# Kapitola 2

## Ranking

### 2.1 Co je to ranking

Veteránský ranking orientačního běhu je ohodnocení závodníků, podle kterého je možné porovnávat jejich výkonnost. Ranking je počítán pro závodníky orientačního běhu starší 35 let. Pro sestavení výsledného pořadí jsou závodníci rozděleni do skupin ve věku  $5n$  až  $5n+4$ , kde  $n \geq 7$ . Tyto skupiny odpovídají kategoriím ve kterých jsou vypisovány závody orientačního běhu.

Ranking je vhodné počítat pro všechny věkové skupiny dohromady, navíc by se měl počítat i ranking všem závodníkům běžajícím v hlavních kategoriích závodů. Důvod je ten, že se závodníci mohou přihlašovat do různých kategorií závodu. Průměr času i rankingu pro výpočet bodů přidělených závodníkům za jednotlivé závody se počítá zvlášť pro každou kategorii závodu. Proto je potřeba umět porovnávat rankingy všem závodníkům běžajícím od hlavní kategorie výše. Nejjednodušší cesta, jak toho docílit, je počítat společný ranking všem závodníkům běžajícím v hlavních a veteránských kategoriích závodů. Jediné rozdělení je na mužský a ženský ranking. Pro výpočet výsledného pořadí se závodníci rozdělí podle věkových skupin, do kterých náleží. V rámci každé skupiny se sestaví pořadí závodníků podle hodnot jejich rankingu.

Je vhodné umožnit počítat dvě varianty rankingu současně, protože se hodnoty rankingu používají ke dvěma různým účelům. První použití je pro určení kvality závodníků, kteří se závodu zúčastnili, při výpočtu bodů přidělených závodníkům za závod. Druhé použití rankingu je pro sestavování pořadí závodníků. Ne nutně musí být oba tyto rankingy stejné, na každý jsou kladeny jiné požadavky. Ranking pro výpočet bodů za závod by měl být stabilní a vypovídat o výkonnosti závodníka, nevádí mírné výchyly na obě strany. Ranking pro výpočet pořadí závodníků by neměl žádné závodníky zvyhodňovat, v odůvodněných případech může některé závodníky znevýhodnit, například za nízký počet odběhnutých závodů.

Ranking závodníka je počítán na základě bodů přidělených závodníkovi za jednotlivé závody. Je ho možné počítat podle nějakého jednoduchého vzorečku, například součet nebo průměr několika nejvyšších bodů přidělených závodníkovi za závod. Dají se použít různé další volby, například vynechat několik nejlepších a nejhorších výsledků závodníka nebo stanovit minimální počet závodů pro zařazení do rankingu. Délka období, ze kterého se závody započítávají, se obvykle stanoví na 1 rok.

Za odběhnutí závodu získává závodník body, z kterých mu bude počítán ranking. Body získané za závod mohou být počítány podle libovolného vzorečku, ve kterém může být jako proměnná použit čas závodníka, průměr a směrodatná odchylka času nebo rankingu rozumně zvoleného počtu nejlepších závodníků v kategorii.

V rámci jedné kategorie závodu lze jako proměnnou chápat pouze čas závodníka, případně jeho umístění. Ostatní parametry jsou v jedné kategorii závodu konstantní. Závodník s nejlepším časem by měl dostat nejvíce bodů přidělených v rámci jeho kategorie a závodník s nejhorším časem by měl dostat bodů nejméně, ostatní závodníci by měli dostat body nějak vhodně odstupňované.



## 2.2 Jak spočítat ranking

Pro spuštění výpočtu rankingů je třeba nejprve vymyslet průběh inicializačního ohodnocení závodníků. Je vhodné přidělit závodníkům nějaký ranking, aby mělo smysl počítat průměrný ranking závodníků v kategorii závodu. Je několik možností jak to mohu udělat. První možnost je přidělit všem závodníkům stejný počet bodů, druhá možnost je použít ranking spočítaný podle jiného vzorečku nebo podle předchozího výpočtu rankingů podle současného vzorečku.

Pro správný běh výpočtu rankingů ještě potřebuji spočítat body přidělené závodníkům za závody z dostatečně dlouhého období. Pokud ranking počítám na základě bodů přidělených za závody za poslední rok, je třeba předpočítat body přidělené závodníkům za závody za alespoň jeden rok dozadu. Teprve poté mohu začít počítat ranking závodníků. V prvním kroku použiji pouze body získané za předpočítané závody, postupem času budu používat stále více bodů ze závodů získaných správným postupem. Po výpočtu rankingů za alespoň jeden rok již nebudu potřebovat žádné z těchto předpočítaných hodnot.

Poté je již vše připraveno a mohu spustit plnohodnotný výpočet rankingů a bodů přidělených za závod. Výpočet se opakuje v měsíčních cyklech. Nejprve je třeba spočítat body přidělené závodníkům za závody, které se konaly během posledního měsíce. Na základě bodů přidělených závodníkům za poslední rok se spočítá ranking závodníků pro následující měsíc. Hodnoty tohoto nového rankingů se použijí pro výpočet průměru rankingů závodníků potřebného pro výpočet bodů za závody následujícího měsíce.

Výpočet bodů přidělených závodníkům za závod se provádí ve dvou krocích. V prvním kroku se spočte počet závodníků v kategorii závodu, průměrný čas, rozptyl času, průměrný ranking a rozptyl rankingů závodníků v kategorii. V druhém kroku se již mohou pro každého závodníka na základě jeho času a umístění, s použitím předchozích údajů, vypočítat body přidělené závodníkovi za tento závod.

Na začátku každého měsíce se ranking aktualizuje. Tento výpočet pobíhá také ve dvou krocích. Nejprve je potřeba aktualizovat pomocný ranking závodníků. Stačí spočítat novou hodnotu jen těm závodníkům, kterým se ranking změnil. To jsou závodníci, kteří běželi v posledním měsíci nějaký závod, nebo závodníci, kterým vypršela platnost nějakého závodu. Pomocný ranking závodníka se spočítá jako průměr nebo součet vzorečkem předepsaného počtu nejlepších závodů za předepsané období.

V druhém kroku výpočtu rankingů se pomocný ranking znormuje tak, aby nejlepší závodník získal 10 000 bodů. Tím získám výsledný ranking pro aktuální měsíc. Pořadí závodníků v jednotlivých veteránských kategoriích se spočítá podle toho rankingů, který je určen pro sestavování výsledkové listiny.

Při aktualizaci rankingů je třeba změnit datum ke kterému se ranking vztahuje.

Pak je již vše připraveno k výpočtu bodů přidělených za závody v následujícím měsíci a poté pro výpočet nového rankingů. Tento postup mohu dále opakovat a tím získám rankingy pro další období.

# Kapitola 3

## Data - výsledky závodů

### 3.1 Zdroj dat

Pro výpočet rankingu jsem potřebovala získat data, tedy výsledky závodů orientačního běhu za několik let zpět. Bohužel neexistuje žádná souhrnná databáze výsledků závodů ve veteránských kategoriích. Výsledky závodů jsem tedy stahovala z různých zdrojů. Povedlo se mi nalézt většinu výsledků závodů za poslední 4 roky, od roku 2004 do roku 2008.

V některých oblastech a krajích měli archiv závodů vedený centrálně, tam nebyl problém výsledky závodů stáhnout. Ve většině oblastí nic takového neměli, proto jsem se snažila data získat ze stránek jednotlivých oddílů orientačního běhu dané oblasti. Tyto stránky byly značně rozdílné kvality. Některé oddíly měly výsledky závodů uložené například v archivu závodů, jiné oddíly měly na svých stránkách výsledky závodů schované nebo měly odkazy, které směřovaly na výsledky závodů, ale tyto stránky již neexistovaly.

Vícedenní závody, které pořádá jeden oddíl opakovaně každý rok pod stejným názvem, mívají své vlastní internetové stránky. Zde jsou uloženy výsledky z posledního roku a většinou i několik let zpět. Mistrovství a Veteraniádu ČR pořádají různé oddíly a zřejmě proto jsem nikde nenašla souhrnný přehled těchto závodů a jejich celkových výsledků.

Výsledky závodů jsem hledala také pomocí Googlu. Zadávala jsem různá klíčová slova, která se ve výsledkových listinách často opakují, jako například výsledek, závod, orientační běh, kontrola a další, tím jsem se snažila nalézt vhodné odkazy. Když jsem našla výsledky nějakého závodu, tak se mi většinou povedlo nalézt ještě několik dalších výsledků závodů úpravou internetové adresy.

Nalezené stránky s výsledky závodů jsem ukládala. Ke každému závodu jsem též potřebovala znát datum, kdy se konal. Tuto informaci jsem zakódovala do názvu souboru. Soubory jsem pojmenovala RRMMDDIII nebo RRMMDD, kde R je rok, M měsíc a D den, kdy se závod konal, I je identifikační číslo závodu. Identifikační čísla závodů byla zvolena Českým svazem orientačního běhu jednoznačně v rámci jednoho roku. Pokud jsem tato čísla nenašla, použila jsem kratší variantu názvu souboru a jednoznačná identifikační čísla později vygeneruji vlastní.

Tímto způsobem jsem získala množství souborů obsahujících výsledky závodů. Potřebovala jsem dekodovat tyto soubory, většinou zapsané v html kódu, a získat výsledky závodů ve formě vhodné pro další zpracování. Potřebovala jsem získat seznam hlavních a veteránských kategorií, pro každou z těchto kategorií seznam závodníků, kteří se závodu zúčastnili a jejich výsledný čas a pořadí.

## 3.2 Formát výsledků závodů

Stránky s výsledky závodů jsou v mnoha různých formátech, dva nejčastější jsou v následujících ukázkách, ale i v těchto formátech je velká variabilita, výsledky mohou být různě formátovány, sloupečky dat přeházené. Nejtěžší je najít a určit kategorii závodu. Označení kategorie bývá na některém řádku poměrně těsně předcházejícím výsledkům dané kategorie. Na řádku s názvem kategorie bývá některé z výrazů "kategorie", "kontrol" spolu s číselným údajem udávající počet kontrol nebo výraz typu "4.3km 50m 13k" udávající délku tratě, převýšení a počet kontrol.

Označení hlavní nebo veteránské kategorie má ve vyjádření pomocí regulárního výrazu tvar `[DH][2-9][105][A-Z]?`, kde DH značí ženskou nebo mužskou kategorii, číslo omezuje věk závodníků, volitelný poslední znak umožňuje vypsát více kategorií pro stejná věková omezení.

Registrační číslo se dá použít jako jednoznačný identifikátor závodníka, stačí tedy z výsledkových listin získávat pouze tento údaj, jméno závodníka již není potřeba. Regulární výraz `[A-Z]{3}[0-9]{4}` popisuje registrační číslo závodníka, kde první tři znaky udávají zkratku oddílu, ve kterém je závodník registrován, následující dvě čísla značí ročník narození, poslední dvojčíslí kódují pohlaví, muži jsou označeni číslem menším než 50, ženy větším než 50, navíc toto dvojčíslí zajišťuje jednoznačnou identifikaci závodníka.

Do výsledků jsem zařadila jen registrované závodníky, kteří úspěšně dokončili závod. Také jsem zkontrolovala, jestli závodníci měli právo startovat v dané kategorii.

Čas závodníka, který bývá většinou uveden v minutách a sekundách, vyjíměčně také v hodinách, minutách a sekundách, jsem pro snazší pozdější zpracování převedla na sekundy.

Dekódované výsledky by měly obsahovat kategorii závodu, registrační číslo závodníka, výsledný čas a umístění. Pro názornost ukázky formátu výsledků závodů

```
<B>Konečné výsledky kategorie: D35          4.3km    0m 13k  </B>
<B>-----</B>
1.Ash Hana                                OCB7051    54.00 +    0.00
2.Hudečková Helena                       KTA6951    54.03 +    0.03
3.Kovaříčková Darja                      KRE6352    57.06 +    3.06
4.Troupová Vladěna                      KRE6351    64.45 +   10.45
5.Stluková Eliška                        PCB6851    68.48 +   14.48
6.Stachová Alena                         VCB6370    73.29 +   19.29
7.Machová Ivana                          KRE6451    76.43 +   22.43
8.Zemanová Alena                         KTA6351    79.56 +   25.56
9.Pešlová Martina                        LEN6751    80.40 +   26.40
. Pecková Ivana                          Ekonom     EKP6252    DISK

Vytvořeno 29.04.2007 21:56
<B>28.04.2007 ŽB Čechy Kunějov - sobota, klasická trať</B>
<B><big><big>D50B</big></big></B>          3.10 km, 110 m, 11 kontrol
<a href="protokol.htm">protokol</a>&nbsp;  &nbsp; <a href="d60b.htm">mezičasy</a>
-----
1. Křtěnská Leontina          BOR6650 C OK Jiskra Nový Bor          47.22
2. Kubíková Květa            LPU5752 C OK Lokomotiva Pardubice 49.18 + 1.56
3. Zmeková Míla              SLP6451 C SK Slavia Praha        50.24 + 3.02
4. Matějů Ladislava         UVP5750 C Slovan Praha          51.29 + 4.07
5. Šlězrová Eva              TUR6350 C OOB TJ Turnov         57.43 +10.21
6. Kovářová Milena          TJH6350 C TJH Jindřichův Hradec 59.07 +11.45
7. Folprechtová Marie       TUR6250 C OOB TJ Turnov         65.54 +18.32
8. Ryšavá Lubomíra          TJP6750 C OOB TJ Sokol Troja    67.31 +20.09
9. Boučková Ludmila         OCB6351 C Orienteering club České Budějo 71.20 +23.58
10.Hudečková Alena          KTA6351 C SK Kotnov Tábor       94.53 +47.31
-----
```

Dalším důležitým formátem výsledků závodů je “povinný formát“. Jde o formát ve kterém jsou pořadatelé závodu povinni zasílat výsledky hlavních kategorií hodnotitelům současného českého rankingu. Jedná se o odlišný formát, zato však s přesně definovanou strukturou. Skript na zpracování výsledků závodů samozřejmě umí zpracovat i tento formát.

Výsledky v “povinném formátu“ mají následující strukturu: kategorie - 10 znaků, jméno závodníka - 25 znaků, registrační číslo závodníka - 7 znaků, licence závodníka - 1 znak, výsledek - 6 znaků (44-49). Výsledek je číslo ve tvaru MMM:SS, kde M značí minuty, S sekundy, vodící 0 není nutná a oddělovač mezi minutami a vteřinami může být libovolný. Pro diskvalifikaci i vzdání je použito číslo 888.88. Výsledky mají být pro přehlednost seřazeny podle kategorií a pořadí v kategorii.

Ukázka formátu dat v “povinném formátu“

H21	Korouhvička Igor	AXC7105A 89.30
H21	Langerová Anastázie	AXB6858B100.30
H21	Špidlovec Lumír	DKD7849B101.40
H21	Loskuták Dmitrij	QWR7845M130.20
H21	Škromach Martin	WIN5902A888.88

### 3.3 Zpracování výsledků závodů

Nejprve jsem výsledky všech závodů, které jsem stáhla, upravila do jednoduchého jednotného formátu. Našla jsem kategorii závodu, registrační číslo závodníka, jeho čas a umístění, tyto hodnoty oddělila čárkami a uložila.

Závody jsem stahovala poměrně zvláštním způsobem, snažila jsem se stáhnout co nejvíce závodů. Tím se stalo, že jsem našla a uložila výsledky několika závodů vícekrát. Musela jsem tedy odstranit tyto duplicity. U některých závodů jsem neznala přesné datum konání závodu, požila jsem například datum zveřejnění výsledků, které bývá uvedeno na výsledkových listinách. Toto datum se příliš neliší od datumu pořádání závodu, pro výpočet rankingu tedy plně dostačuje. Pro nalezení duplicitních výsledků závodu je potřeba zvolit metodu, která není založená na datech pořádání závodu. Rozhodla jsem se porovnávat dekódované výsledky závodu. Pokud jsem našla dva soubory s výsledky závodu, kde běželi stejní závodníci ve stejných kategoriích, doběhli ve stejném pořadí a ve stejném čase, pak jde zřejmě o výsledky stejného závodu.

Dále bylo potřeba přidělit jednotlivým závodům jednoznačné identifikační číslo, například identifikační číslo závodu tvaru RMMDDIII, kde R je rok, M měsíc a D den konání závodu, I libovolné, tak aby zajišťovalo jednoznačnost identifikačního čísla. Pro některé závody jsem znala identifikační číslo určené Českým svazem orientačního běhu, pak jsem zkontrolovala jednoznačnost a zvolila toto. Český svaz orientačního běhu určuje identifikační číslo závodu jako datum konání závodu a pořadové číslo závodu v daném roce, každý rok se pořádá kolem 200 závodů. V případech, kdy jsem neznala správné identifikační číslo, jsem pro poslední trojčíslí zvolila číslo vyšší než 300 tak, aby celé identifikační číslo závodu bylo jednoznačné.

Takto zpracované výsledky závodů jsem uložila do databáze.

Nové výsledky závodů lze snadno zpracovat a vložit do databáze pomocí těchto skriptů. Stačí uložit výsledky do správného adresáře a spustit zpracování. Skript sám pozná soubory obsahující nové výsledky závodů podle toho, že neexistuje soubor stejného názvu se zpracovanými výsledky. Skript dekóduje výsledky závodu, provede kontrolu, zda se nejedná o duplicitní výsledek, tedy jestli je výsledek závodu odlišný od všech ostatních, přidělí závodu jednoznačné identifikační číslo a uloží data do databáze.

# Kapitola 4

## Struktura databáze

Databáze se skládá z několika logických celků. Jedna část databáze popisuje závodníky a výsledky závodů bez ohledu na vzoreček pro výpočet rankingu, další přiděluje ranking závodníkům a body získané za závod v závislosti na zvoleném vzorečku. V databázi jsou také zakódovány vzorečky pro výpočet rankingu a bodů přidělených za závody i se svými parametry.

### 4.1 Závodníci a výsledky závodů

Část databáze zobrazující závodníky a výsledky závodů bez ohledu na použitý vzoreček obsahuje tabulky *Závodníci*, *Oddíly*, *Závody* a *Výsledky*. Tato část databáze obsahuje data představující vstup pro výpočet rankingu.

Tabulka *Závodníci* popisuje všechny závodníky registrované v Českém svazu orientačního běhu. Pro každého závodníka je zde uvedeno jméno, registrační číslo, které slouží jako primární klíč tabulky, rok narození závodníka, oddíl orientačního běhu, kde je závodník registrovaný, a veteránská věková kategorie, do které závodník spadá, pro příliš mladé závodníky zůstává tato položka nedefinovaná. Údaje o závodnících jsem získala ze seznamu registrovaných závodníků Českého svazu orientačního běhu.

Tabulka *Oddíly* obsahuje názvy a zkratky všech oddílů orientačního běhu. Jako primární klíč slouží zkratka oddílu. Tyto údaje jsem získala z adresáře Českého svazu orientačního běhu.

Tabulka *Závody* obsahuje údaje o minulých i plánovaných závodech orientačního běhu. Tabulka obsahuje pro každý závod identifikační číslo závodu, sloužící jako primární klíč, koeficient závodu, který vyjadřuje důležitost závodu a může být použit pro výpočet rankingu a typ startu závodu. K závodům pořádaným od roku 2007 až dosud jsou zde navíc uloženy podrobnější informace o závodu, tedy oddíl pořádající závod, název závodu, oblast do jehož žebříčku závod spadá, místo, kde se závod konal, a název mapy, na které se závod běžel. Tyto údaje jsem získala z kalendáře závodů Českého svazu orientačního běhu ze sekce stažení dat.

Tabulka *Výsledky* obsahuje výsledky závodů, tedy identifikační číslo závodu a registrační číslo závodníka, sloužící jako primární klíč, kategorii závodu, ve které závodník běžel, čas a umístění kterého závodník dosáhl. V této tabulce jsou uloženy výsledky většiny závodů orientačního běhu, které se konaly za poslední čtyři roky. Postup získání těchto údajů je popsán v předchozí kapitole.

## 4.2 Ranking a body za závod

Další část databáze umožňuje počítat ranking závodníků a body přidělené závodníkům za závody podle různých vzorečků. Obsahuje tabulky Kategorie, Výsledky Body, Ranking, Ranking Pomocný, Nejnovější Závod, Stav Databáze a Archiv. Všechny tyto tabulky obsahují atribut Vzoreček, který udává číslo použitého vzorce, jako část primárního klíče. Tím je umožněno použití stejných tabulek pro výpočet rankingu podle různých vzorců.

Tabulka *Kategorie* obsahuje pro každou kategorii závodu počet závodníků, kteří mají přidělen ranking pro výpočet bodů za závod a také průměry a rozptyly času, průměry a rozptyly rankingu vzorečkem předepsaného počtu nejlepších závodníků. Primární klíč tabulky tvoří číslo závodu, kategorie závodu a použitý vzoreček.

Tabulka *Výsledky Body* obsahuje body, které byly přidělené závodníkům za jednotlivé závody. Dále jsou zde hodnoty obou typů rankingů, tedy rankingu pro výpočet bodů za závod i rankingu pro tvorbu výsledkových listin. Hodnoty těchto rankingů odpovídají hodnotám, které byly pro závodníka aktuální v době konání závodu. Primární klíč této tabulky tvoří číslo závodu, registrační číslo závodníka a použitý vzoreček.

Do tabulky *Nejnovější Závod* se před výpočtem pomocného rankingu zapíše číslo závodu, který závodník běžel naposled. Primární klíč této tabulky tvoří registrační číslo závodníka a číslo použitého vzorečku. Pomocí data posledního závodu jednotlivých závodníků se snadno pozná, jestli během posledního měsíce běželi závod, tedy je třeba přepočítat jejich ranking nebo ne. Tento údaj by se měl dát snadno spočítat pomocí vnořeného dotazu. Používala jsem databázi na Malé Straně, kde je nainstalována MySQL verze 4.0.18, která bohužel ještě neumí zpracovat vnořené dotazy a proto jsem byla nucena vytvořit tuto tabulku.

Tabulka *Ranking Pomocný* slouží k uložení výsledků prvního kroku výpočtu rankingu. Primární klíč tabulky tvoří registrační číslo závodníka a typ uloženého rankingu a použitý vzoreček. Dále je v tabulce uložena hodnota pomocného rankingu, počet závodů, jejichž body přidělené za závod byly použity při výpočtu rankingu, číslo nejstaršího z těchto závodů, aby se bylo možno zkontrolovat, jestli je třeba ranking přepočítat kvůli novému závod, a nejnižší z bodů přidělených za závod a použitých při výpočtu pomocného rankingu.

Tabulka *Ranking* je důležitá tabulka, která obsahuje aktuální ranking závodníků spočítaný podle různých vzorečků. Obsahuje aktuální hodnotu rankingu pro výpočet bodů přidělených za závody, aktuální hodnotu rankingu pro tvorbu výsledkových listin a pořadí závodníků v rámci jednotlivých veteránských kategorií. Primární klíč této tabulky tvoří registrační číslo závodníka a číslo použitého vzorečku.

Tabulka *Stav Databáze* umožňuje uložit pro každý vzoreček datum, ke kterému je aktuální ranking spočítaný podle tohoto vzorečku a uložený v tabulce Ranking.

Tabulka *Archiv* slouží k ukládání starých hodnot rankingu závodníků. V tabulce Ranking se hodnota rankingu přepisuje za hodnotu aktuálního rankingu a zároveň se stará hodnota rankingu zálohuje do tabulky Archiv. Primární klíč této tabulky tvoří registrační číslo závodníka, datum, ke kterému byl ranking aktuální a číslo vzorečku podle kterého byl ranking spočítán.

## 4.3 Vzorečky v databázi

Poslední část databáze popisuje samotné vzorečky pro výpočet rankingů závodníků a bodů přidělených za závody. Tabulka *Vzoreček Obecně* obsahuje souhrnné informace o vzorečkách. Je zde uloženo číslo vzorečku, jeho název a slovní popis.

Tabulka *Vzoreček Závod* popisuje jak z výsledků závodů spočítat body získané za závod. Primární klíč tabulky tvoří číslo vzorečku, kategorie závodu a typ startu. Pro různé kategorie a typy startu umožňuje zadat různé vzorečky pro výpočet bodů. V tabulce jsou uloženy následující parametry popisující chování vzorečku

- minimální počet závodníků v kategorii potřebný pro hodnocení této kategorie,
- počet časově nejlepších a rankingově nejlepších závodníků, jejichž výsledky se použijí pro výpočet průměru a rozptylu času a rankingů závodníků kategorie,
- možnost zvolit tento počet jako pevné číslo nebo číslo relativní k počtu závodníků v kategorii,
- způsob úpravy hodnoty času a rankingů před použitím při výpočtu
- minimální počet bodů, které závodník získá za úspěšně dokončený závod.

Samotný vzoreček pro výpočet bodů přidělených závodníkovi za závod je zadán formulí, která je také zapsána v této tabulce. Formule pro výpočet je zapsána pomocí PHP kódu, který má za úkol vypočítat přidělený počet bodů a uložit jej do proměnné pojmenované *\$body*. Během výpočtu může používat následující proměnné

<i>\$cas</i>	představující čas závodníka,
<i>\$umisteni</i>	představující umístění závodníka v závodě v rámci kategorie,
<i>\$ranking</i>	představující ranking závodníka v čase závodu,
<i>\$prumcas</i>	představující průměrný čas zvoleného počtu časově nejlepších závodníků,
<i>\$varcas</i>	představující rozptyl časů zvoleného počtu časově nejlepších závodníků,
<i>\$prumrank</i>	představující průměrný ranking zvoleného počtu závodníků s nejvyšším rankingem sloužícím k výpočtu bodů za závody,
<i>\$varrank</i>	představující rozptyl rankingů zvoleného počtu závodníků s nejvyšším rankingem sloužícím k výpočtu bodů za závody,
<i>\$pocet</i>	představující počet závodníků v kategorii,
<i>\$koeficient</i>	představující koeficient závodu určený Českým svazem orientačního běhu.

Tabulka *Vzoreček Výsledky* popisuje postup, jak se z bodů získaných závodníkem za jednotlivé závody spočítá ranking závodníka. Primární klíč této tabulky tvoří číslo vzorečku, věková kategorie závodníka a typ rankingů. Je možné zároveň počítat ranking, který se použije při výpočtu bodů přidělených závodníkům za závody, a ranking, podle kterého se sestaví výsledné pořadí závodníků v rámci jednotlivých veteránských kategorií. Parametry pro výpočet rankingů závodníků, které je možno zvolit jsou následující

- počet nejlepších závodů jejichž výsledky se zohledňují při výpočtu rankingů,
- počet nejlepších a nejhorsích závodů, které se nezapočítají,
- minimální počet závodů, aby závodníkovi byl přidělen ranking,
- délka období z kterého se vybírají závody, například nejvýše rok staré závody,
- agregační funkce podle které se počítá ranking, například průměr nebo součet.

Pomocí těchto tabulek je možné jednoduše vytvořit téměř libovolný vzoreček, jaký si vymyslíte. Systém umí spočítat body přidělené závodníkům za závody, ranking jednotlivých závodníků a sestaví pořadí závodníků ve veteránských kategoriích podle Vámi zvoleného vzorečku.

# Kapitola 5

## Uživatelské rozhraní

### 5.1 Stránky zobrazující ranking

Uživatelské rozhraní je tvořeno HTML stránkami generovanými pomocí PHP skriptů. Struktura stránek je vzájemně provázána pomocí menu a pomocí odkazů v rámci samotných stránek. Je zde zobrazen ranking závodníků a výsledky závodů. Uživatelé mohou prohlížet jednotlivé stránky a pohybovat se po nich pomocí odkazů.

Stránka zobrazující ranking závodníků umožňuje prohlížení výsledného rankingu ve všech veteránských věkových kategoriích. Výpis závodníků lze setřídít podle pořadí v kategorii, abecedně podle jména závodníka nebo podle oddílu. Pro každého závodníka je zobrazeno jeho pořadí v kategorii, jméno, registrační číslo, ranking sloužící pro určení pozice ve výsledkové listině a ranking používaný pro výpočet bodů získaných za jednotlivé závody. Jméno závodníka slouží jako odkaz na podrobnější informace o rankingu daného závodníka.

Stránka zobrazující podrobnější informace o rankingu je společná pro závodníky stejného oddílu. Pro každého závodníka je zde seznam závodů, které běžel během posledního roku, tedy seznam závodů podle kterých se počítá jeho ranking. K jednotlivým závodům je vypsáno datum konání, pořádající oddíl, kategorie, ve které závodník běžel, získané body a umístění v závodě. Odkazy vedou na stránky výsledků závodů.

Stránka výsledků závodu obsahuje podrobnější informace o závodě a samozřejmě také výsledky závodu po jednotlivých kategoriích. Informace obsahují název závodu, pořádající oddíl, datum konání závodu, identifikační číslo závodu, koeficient závodu vypovídající o jeho důležitosti, způsob startu, oblast, které se závod započítal do žebříčku, místo, kde se závod konal, a mapu, na které se závod běžel. Následují výsledky jednotlivých kategorií závodu. Pro každou kategorii je vypsán průměr a rozptyl času a rankingu několika nejlepších závodníků v kategorii, který se použil při výpočtu bodů přidělených závodníkům za závod.

Samotné výsledky závodu jsou seřazeny podle umístění závodníků. Pro každého závodníka je vypsáno jeho umístění, jméno, registrační číslo, které slouží jako odkaz na podrobnější informace o rankingu závodníka, veteránská kategorie, pokud jde o veteránského závodníka, čas dosažený v závodě, body získané za závod, ranking používaný na určení výsledného pořadí a ranking použitý při výpočtu bodů za závod. Tyto hodnoty rankingu byly pro závodníka aktuální v době konání závodu.

Stránka obsahující přehled závodů umožňuje zobrazení výsledků konkrétního závodu. Na této stránce je seznam všech závodů pořádaných od roku 2004 a zahrnutých do veteránského rankingu. Pro každý z těchto závodů je zobrazen datum konání, název závodu, pořádající oddíl a oblast, ve které se závod konal.

Stránka se seznamem oddílů orientačního běhu obsahuje zkratky a celé názvy všech oddílů orientačního běhu. Z této stránky vedou odkazy na stránky obsahující podrobnější informace o rankingu závodníků oddílu.

Stránky zobrazující ranking závodníků jsou na adrese <http://www.ms.mff.cuni.cz/~koval6am/ranking.php>, kde si je mohou zájemci prohlédnout.



# Kapitola 6

## Statistika

### 6.1 Základní pojmy

Nechť  $\Omega$  je *prostor elementárních jevů*, nechť  $A$  je  $\sigma$ -algebra podmnožin prostoru  $\Omega$ . Nechť pravděpodobnostní míra  $P$  připisuje pravděpodobnost jednotlivým množinám patřícím do  $A$  tak, že platí

$$P(\emptyset) = 0, P(\Omega) = 1$$

$$\forall A \in A \quad P(A) \in [0,1]$$

$$\forall \{A_i\}_{i=1}^{\infty} \subset A, A_i \cap A_j = \emptyset \text{ pro } i \neq j \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Pak se trojice  $(\Omega, A, P)$  nazývá *pravděpodobnostní prostor*.

Nechť  $R$  je množina reálných čísel a  $B$  systém jejích borelovských podmnožin. Nechť  $X(\omega)$  je měřitelná funkce z  $(\Omega, A, P)$  do  $(R, B)$ . Pak se  $X(\omega)$  nazývá *náhodná veličina* a značí se stručně  $X$ . Ke každé borelovské množině  $B \in B$  lze vyjádřit její vzor  $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\}$  a pravděpodobnostní míru  $Q(B) = P(X^{-1}(B))$ . Míra  $Q$  se nazývá *zákon rozdělení náhodné veličiny*, většinou se říká stručně jen *rozdělení*.

Pro  $B = (-\infty, x)$  definujeme *distribuční funkci*  $F$  předpisem  $F(x) = P(X < x)$ .

Distribuční funkce je neklesající, zleva spojitá a platí  $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow +\infty} F(x) = 1$ .

Nechť  $X$  je náhodná veličina definovaná na pravděpodobnostním prostoru  $(\Omega, A, P)$ . Pak definujeme *střední hodnotu* náhodné veličiny  $X$  jako

$$EX = \int_{\Omega} X(\omega) dP(\omega)$$

pokud uvedený integrál existuje. Existuje-li  $EX$ , pak platí  $E(aX + b) = aEX + b$ .

Definujme  $k$ -tý *moment* jako  $\mu'_k = EX^k, k = 1, 2, \dots$ . Existuje-li moment  $\mu'_1$  a je-li konečný, definujme  $k$ -tý *centrální moment* předpisem  $\mu_k = E(X - EX)^k, k = 0, 1, 2, \dots$ . Nejdůležitějším centrálním momentem je  $\mu_2$ , který se nazývá *rozptyl* a značí  $\text{var}X$ . Odmocninu z rozptylu  $\sqrt{\text{var}X}$  nazýváme *směrodatná odchylka*.

Existuje-li  $EX$  a je-li  $EX^2 < \infty$ , pak platí  $\text{var}(aX + b) = a^2 \text{var}X$ .

Nechť  $X_1, \dots, X_n$  jsou náhodné veličiny definované na stejném pravděpodobnostním prostoru  $(\Omega, A, P)$ , pak vektor  $X = (X_1, \dots, X_n)$  nazveme *náhodný vektor*.

Náhodné veličiny  $X_1, \dots, X_n$  se nazývají *nezávislé*, pokud platí pro libovolné borelovské množiny vztah  $P\left(\bigcap_{k=1}^n \{\omega : X_k(\omega) \in B_k\}\right) = \prod_{k=1}^n P(\omega : X_k(\omega) \in B_k)$ .

Jestliže  $EX_i^2 < \infty$  pro  $i = 1, \dots, n$ , pak definujeme *kovarianci*  $\text{cov}(X_i, X_j)$  náhodných veličin  $X_i, X_j$  vztahem  $\text{cov}(X_i, X_j) = E(X_i - EX_i)(X_j - EX_j)$

## 6.2 Korelační koeficient

Nechť  $X$  a  $Y$  jsou náhodné veličiny s konečnými druhými momenty a s kladnými rozptyly. Závislost těchto veličin na sobě se často měří pomocí *korelačního koeficientu*

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sqrt{(\text{var } X)(\text{var } Y)}}.$$

Nechť  $a, b, c, d$  jsou taková reálná čísla, že platí  $ac \neq 0$ . Pak platí

$$\rho_{aX+b,cY+d} = \begin{cases} \rho_{X,Y} & \text{pokud } ac > 0 \\ -\rho_{X,Y} & \text{pokud } ac < 0 \end{cases}$$

Pro korelační koeficient platí  $-1 \leq \rho_{X,Y} \leq 1$ .

Rovnost  $\rho_{X,Y} = 1$  platí právě tehdy, když  $Y = aX + b$ , kde  $a > 0$ , obdobně rovnost  $\rho_{X,Y} = -1$  platí právě tehdy, když  $Y = aX + b$ , kde  $a < 0$ .

## 6.3 Výběrový korelační koeficient

Nechť  $X_1, \dots, X_n$  je posloupnost nezávislých stejně rozdělených náhodných veličin s rozdělením  $Q$ . Pak říkáme, že  $X_1, \dots, X_n$  je *náhodný výběr* z rozdělení  $Q$ .

Mějme náhodný výběr  $(X_1, Y_1), \dots, (X_n, Y_n)$  z nějakého dvojrozměrného rozdělení. Definujeme *výběrový průměr*  $\bar{X}$  výběru  $X_1, \dots, X_n$  jako

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

dále definujeme *výběrový rozptyl*  $S_X^2$  výběru  $X_1, \dots, X_n$  jako

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

*výběrovou směrodatnou odchylku*  $S_X$  definujeme jako

$$S_X = \sqrt{S_X^2}.$$

Obdobně  $\bar{Y}$  výběrový průměr a  $S_Y^2$  výběrový rozptyl výběru  $Y_1, \dots, Y_n$ .

Definujeme *výběrovou kovarianci*  $S_{X,Y}$  jako

$$S_{X,Y} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

Je-li  $S_X^2 > 0$  a  $S_Y^2 > 0$ , definujeme *výběrový korelační koeficient*  $r_{X,Y}$  jako

$$r_{X,Y} = \frac{S_{X,Y}}{\sqrt{S_X^2 S_Y^2}}.$$

Pro výběrový korelační koeficient platí  $-1 \leq r_{X,Y} \leq 1$ .

Nechť  $a, b, c, d$  jsou taková reálná čísla, že platí  $ac \neq 0$ . Pak platí

$$r_{aX+b,cY+d} = \begin{cases} r_{X,Y} & \text{pokud } ac > 0 \\ -r_{X,Y} & \text{pokud } ac < 0 \end{cases}$$

## 6.4 Spearmanův korelační koeficient

Hodnocení výběrového korelačního koeficientu je vázáno na splnění předpokladu, že výběr pochází z normálního rozdělení. Dost často se však stává, že tento předpoklad není splněn. Ale pokud jsou pořadí  $X$ -ových a  $Y$ -ových veličin hodně podobná, svědčí to o závislosti mezi  $X_i$  a  $Y_i$ .

Předpokládejme, že  $(X_1, Y_1), \dots, (X_n, Y_n)$  je výběr ze spojitého dvojrozměrného rozdělení. Necht'  $R_1, \dots, R_n$  je pořadí veličin  $X_1, \dots, X_n$  a necht'  $Q_1, \dots, Q_n$  je pořadí veličin  $Y_1, \dots, Y_n$ .

*Spearmanův korelační koeficient*  $r_s$  se definuje jako výběrový korelační koeficient počítaný z dvojic  $(R_1, Q_1), \dots, (R_n, Q_n)$ . Je spočítán jako

$$r_s = \frac{\sum R_i Q_i - n\bar{R}\bar{Q}}{\sqrt{(\sum R_i^2 - n\bar{R}^2)(\sum Q_i^2 - n\bar{Q}^2)}}.$$

## 6.5 Modifikovaný Spearmanův korelační koeficient

Modifikovaný Spearmanův korelační koeficient jsem navrhla pro hodnocení vhodnosti jednotlivých vzorečků pro výpočet rankingu. Ověřuji jestli pořadí závodníků podle času je podobné pořadí závodníků podle rankingu, který měli před závodem. Pokud bych toto chtěla ověřit pro jednu kategorii jednoho závodu, použila bych Spearmanův korelační koeficient, ale závodů je mnoho. Pokud bych spočítala Spearmanův korelační koeficient jen pro několik málo vybraných kategorií několika závodů, vypovídající hodnota těchto výsledků by byla nízká. Pokud bych spočítala Spearmanův korelační koeficient zvlášť pro všechny kategorie všech závodů, dostala bych mnoho korelačních koeficientů a ty bych musela nějak dále zpracovat, například počítat průměr. Potom by ale měly kategorie s málo závodníky stejnou váhu jako kategorie více obsazené, a to nechci, protože pokud například skončí závodník s nejhorším rankingem na prvním místě a závodník s nejlepším rankingem na poslední místě, má tento výsledek úplně jiný význam pro závod se dvěma nebo 20 účastníky.

Pomocí modifikovaného Spearmanova korelačního koeficientu mohu posoudit vhodnost jednotlivých vzorečků navržených pro výpočet rankingu.

Mějme výběr ze spojitého dvojrozměrného rozdělení, který je rozdělený do  $N$  skupin po  $n_1, \dots, n_N$  prvcích, označme je  $(X_{1,1}, Y_{1,1}), \dots, (X_{1,n_1}, Y_{1,n_1}), \dots, (X_{N,1}, Y_{N,1}), \dots, (X_{N,n_N}, Y_{N,n_N})$ . Necht' pro  $i$ -tou skupinu je  $R_{i,1}, \dots, R_{i,n_i}$  pořadí veličin  $X_{i,1}, \dots, X_{i,n_i}$  (např. času), a  $Q_{i,1}, \dots, Q_{i,n_i}$  je pořadí veličin  $Y_{i,1}, \dots, Y_{i,n_i}$  (např. rankingu).

*Modifikovaný Spearmanův korelační koeficient*  $\tilde{r}_s$  potom definuji jako výběrový korelační koeficient počítaný z dvojic  $(R_{1,1}, Q_{1,1}), \dots, (R_{1,n_1}, Q_{1,n_1}), \dots, (R_{N,1}, Q_{N,1}), \dots, (R_{N,n_N}, Q_{N,n_N})$ . Je spočítán jako

$$\tilde{r}_s = \frac{\sum_{i=1}^N \sum_{j=1}^{n_i} R_{i,j} Q_{i,j} - N\bar{R}\bar{Q}}{\sqrt{\left(\sum_{i=1}^N \sum_{j=1}^{n_i} R_{i,j}^2 - N\bar{R}^2\right) \left(\sum_{i=1}^N \sum_{j=1}^{n_i} Q_{i,j}^2 - N\bar{Q}^2\right)}}.$$

# Kapitola 7

## Vzorečky

### 7.1 Vzoreček pro výpočet rankingu

Ranking závodníků je jejich ohodnocení podle kterého je možné porovnat výkonnost závodníků. Ranking lze počítat mnoha způsoby. Je vhodné počítat souběžně dvě varianty rankingu. Jedna varianta může sloužit k určení kvality závodu pomocí rankingu závodníků, kteří se závodu zúčastnili a druhá k sestavení pořadí závodníků. Na každou z těchto variant jsou kladeny mírně jiné požadavky.

Ranking k určení kvality závodníků, tedy ranking sloužící pro výpočet průměru a směrodatné odchylky rankingu nejlepších závodníků v kategorii závodu, by měl pokud možno co nejpřesněji popisovat výkonnost jednotlivých závodníků. Pro výpočet rankingu lze použít průměr několika nejvyšších bodů přidělených závodníkovi za jeho závody. Tento ranking by se měl přidělit pouze těm závodníkům, kteří odběhli alespoň nějaký minimální počet závodů během posledního roku. Předchozí opatření je potřebné pro zajištění stability rankingu. Pokud bych tímto postupem počítala ranking závodníků, kteří mají odběhnuto jen málo závodů, pak by ranking nevypovídal o výkonnosti závodníka, mohlo by se totiž jednat pouze o závody, které se závodníkovi vyjíměčně podařili nebo naopak neobvykle nepovedly, jeho ranking by jimi byl výrazně ovlivněn.

Pro výpočet rankingu pro určení pořadí závodníků ve výsledkové listině se může použít součet bodů přidělených závodníkovi za jednotlivé závody. Tento ranking může být klidně přidělen všem závodníkům, kteří odběhli během posledního roku alespoň jeden závod. Takto spočítaný ranking sice znevýhodní závodníky s nízkým počtem závodů, ale pokud by se tito závodníci chtěli lépe umístit ve veteránském rankingu, stačí jim běhat více závodů.

Rozdílnost mezi vzorečky používajícími průměr nebo součet bodů při výpočtu rankingu nemusí být příliš výrazná. Rankingy spočítané pomocí součtu a průměru bodů z  $x$  nejlepších závodů by vycházely obdobně, pokud by všichni závodníci měli odběhnutých alespoň  $x$  závodů za poslední rok. Rozdíl ve výpočtu se projeví pouze u těch závodníků, kteří mají méně než požadovaných  $x$  závodů. Každý z těchto způsobů výpočtu rankingu se hodí k mírně jinému účelu.

Jako jiná možnost výpočtu rankingu, by se mohl počítat součet bodů ze všech závodů, kterých se závodník zúčastnil za poslední rok. Tento způsob výpočtu ale nemá příliš význam. Někteří závodníci mají odběhnuto výrazně více závodů než ostatní. Ranking spočítaný jako součet všech bodů by je příliš zvýhodňoval.

Ranking počítaný jako průměr všech závodů by naopak mohl zvýhodňovat závodníky, kteří mají málo závodů, které se jim velice povedly. Dalo by se uvažovat o způsobu výpočtu, kdy se závodníkům s nízkým počtem závodů ranking sníží, například pronásobením koeficientem menším než 1, ostatním ranking spočtu jako průměr všech bodů získaných za jednotlivé závody. Tento způsob výpočtu rankingu by ale mohl navádět závodníky k rozhodnutí raději vynechat závod, pokud si budou myslet, že by mohli dostat přidělený nízký počet bodů a tím by se jim snížil průměr jejich bodů a tedy i celkový ranking.

## 7.2 Vzoreček pro výpočet bodů přidělených za závod

Vzoreček pro výpočet bodů za závod může být téměř libovolný výraz. Tento výraz by měl nějakým způsobem vyjadřovat úspěšnost závodníka v závodě, tedy vítězovi přidělit nejvíce bodů a hůře umístěným závodníkům bodů méně, pokud možno odstupňovaně podle rozestupů mezi závodníky. Orientační běh je individuální sport, kde závodníci ze stejné kategorie startují obvykle každý v jiném čase, podobný výsledný čas tedy znamená podobný výkon v závodě, proto by měl vzoreček udělovat závodníkům, kteří doběhli v podobném čase, podobný počet bodů.

Výpočet bodů přidělených za závod by měl záviset na čase závodníka a porovnat tento čas s časy ostatních závodníků v kategorii závodu. Případně se může použít umístění závodníka, nejspíše v poměru s celkovým počtem závodníků v kategorii. Kvalitu závodníků v kategorii závodu je možné vyjádřit pomocí průměru a směrodatné odchylky rankingu závodníků, kteří se závodů zúčastnili.

Používala jsem dva základní typy vzorečku v několika variantách.

$$B_z = 1000 - 200 * \frac{\check{C}_z - \check{C}_p}{\check{C}_s} \qquad B_z = 1000 * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$$

$$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s} \qquad B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$$

$$B_z = 1000 - 200 * \frac{\log(\check{C}_z) - \check{C}_{p_{\log}}}{\check{C}_{s_{\log}}} \qquad B_z = 1000 * \left( 2 - \frac{\log(\check{C}_z)}{\check{C}_{p_{\log}}} \right)$$

- kde  $B_z$  jsou body které závodník získal v závodě,  
 $\check{C}_z$  je čas závodníka,  
 $\check{C}_p$  je průměrný čas, určený z časů nejrychlejších závodníků, kteří mají v době závodu přidělený ranking pro výpočet bodů za závod  
 $\check{C}_s$  je směrodatná odchylka časů, určená z časů nejrychlejších závodníků, kteří mají v době závodu přidělený ranking pro výpočet bodů za závod  
 $R_p$  je průměrný ranking rankingově nejlépe ohodnocených závodníků  
 $R_s$  je směrodatná odchylka rankingu rankingově nejlépe ohodnocených závodníků  
 $\check{C}_{p_{\log}}$  je průměr spočítaný z logaritmu času nejrychlejších závodníků, kteří mají v době závodu přidělený ranking pro výpočet bodů za závod  
 $\check{C}_{s_{\log}}$  je směrodatná odchylka spočítaná z logaritmu času nejrychlejších závodníků, kteří mají v době závodu přidělený ranking pro výpočet bodů za závod

Ve vzorečkách, kde se při výpočtu bodů za závod používá ranking závodníků je vhodné používat stejný počet časově nejlepších závodníků pro časové průměry jako je použito rankingově nejlepších závodníků pro rankingové průměry. V ideálním případě, kdy by závodníci doběhli v pořadí od rankingově nejlepšího po rankingově nejhoršího, by se časové a rankingové průměry a směrodatné odchylky počítaly z časů a z rankingů stejných závodníků.

## 7.3 Parametrizace vzorečku

Důležitý parametr, který je nutno zvolit, je počet závodníků, z jejichž výsledků se budou počítat průměry a směrodatné odchylky časů a rankingů, které se dále použijí pro výpočet bodů získaných závodníky za jednotlivé závody.

První varianta je vybrat nějaké pevně zvolené, poměrně nízké číslo. Za touto volbou stojí snaha porovnávat časy závodníků s časem vítěze. Použití přímo času vítěze ubírá na stabilitě výpočtu, i vítězovi se může závod povést lépe nebo hůře. Pokud se použije průměrný čas několika nejlepších závodníků, bude tento čas o trochu nižší, ale zato s větší vypovídající hodnotou.

Druhá varianta je počítat useknutý průměr a směrodatnou odchylku času nebo rankingů z časů a rankingů relativního počtu nejlepších závodníků. Například pro 90% průměry a směrodatné odchylky času a rankingů bych pro závod s 10 závodníky bych počítala průměrný čas z časů 9 nejrychlejších závodníků a průměrný ranking z rankingů 9 závodníků s nejvyšším rankingem. Pokud by se počítaly průměry a směrodatné odchylky z časů a rankingů všech závodníků, byly by výsledné hodnoty příliš ovlivněny časy posledních závodníků, kterým se závod mohl výrazně nepovést, proto je lepší několik výsledků nejhorších závodníků nezapočítat.

V první variantě výpočtu dostanou závodníci, kteří se umístili na prvních místech, vždy podobný počet bodů a body dalších závodníků budou odstupňovány podle ztráty na tyto první závodníky. Druhá varianta vychází trochu jinak. Podobný počet bodů dostanou závodníci ze středu výsledkové listiny. Vítězové mohou získat značně rozdílný počet bodů podle času o který se jim podaří předběhnout většinu závodníků.

Další rozdíl mezi těmito variantami výpočtu se objeví pro závody s výrazně jiným počtem závodníků. Pokud pro každý závod počítám průměry a směrodatné odchylky času nebo rankingů například z hodnot 4 nejlepších závodníků, má tento průměr jiný význam pro závod kde jsou 4 závodníci, to bych počítala pro průměr hodnot všech závodníků, a jiný význam pro závod, kde je například 20 závodníků, zde by to byl průměr hodnot pětiny nejlepších závodníků.

## 7.4 Motivace pro různé vzorečky

Nyní bych chtěla napsat motivaci k jednotlivým vzorečkům. Nejprve vzoreček

$$B_z = 1000 - 200 * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$$

Vzoreček je odvozen z předpokladu, že časy závodníků v kategorii závodu mají normální rozdělení. Potom také body přidělené závodníkům za závod budou mít normální rozdělení, které bude mít střední hodnotou 1000 a rozptyl 200.

Další vzoreček, který se snaží vylepšit předcházející, vypadá následovně

$$B_z = 1000 * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$$

Tento vzoreček mohu upravit následujícím způsobem, tím získám podobný tvar výrazu jako v předchozím vzorečku.

$$B_z = 1000 * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right) = 1000 * \left( 1 - \frac{\check{C}_z - \check{C}_p}{\check{C}_p} \right) = 1000 - 200 * \frac{\check{C}_z - \check{C}_p}{0,2 \check{C}_p}$$

Hodnotu  $0,2 * \check{C}p$ , což znamená 0,2 krát průměrný čas závodníků v kategorii, můžeme považovat za odhad směrodatné odchylky času závodníků v kategorii závodu. Tento předpoklad není úplně nesmyslný, protože mezi průměrným časem a směrodatnou odchylkou časů je poměrně silná korelace, jak je ukázáno v následující kapitole. Tento vzoreček tedy můžeme považovat za odhad prvního vzorečku.

První vzoreček není příliš vhodný, pokud je v kategorii málo závodníků, a to se stává poměrně často. Výběrová směrodatná odchylka vychází značně odlišně od teoretické hodnoty, kterou by měl závod mít. Pokud se stane, že všichni (například dva) závodníci doběhnou v podobném čase, a to se může stát a také stalo, pak bude směrodatná odchylka velice malá, tedy první závodník dostane velmi mnoho bodů a druhý jen velice málo, přitom jejich výsledek v závodě byl téměř stejný. Pokud by závodníci doběhli s velkým časovým rozestupem, bude směrodatná odchylka velká a závodníkům bude přidělen podobný počet bodů a přitom jejich výsledek v závodě byl značně rozdílný. Tyto nedostatky druhý vzoreček nemá.

Závodů se koná velké množství různé obtížnosti. Pokud se například závodu zúčastní převážně běžci s vysokým rankingem, tedy dobří závodníci, bude průměrný čas závodu rychlý, pokud by se stejného závodu zúčastnili běžci s nízkým rankingem, tedy méně schopní závodníci, byl by průměrný čas daleko nižší, tedy horší. Závodník, který by se zúčastnil obou těchto závodů a doběhl ve stejném čase, by pak v prvním závodě skončil na konci a dostal málo bodů, v druhém závodě by se stejným časem mohl zvítězit a dostat hodně bodů, přitom jeho výkon byl v obou případech stejný. Je třeba nějakým způsobem rozlišit různou úroveň závodů.

Můžeme přidat do vzorečku průměr a směrodatnou odchylku rankingu závodníků, kteří se zúčastnili závodu. Použije se ranking určený k výpočtu bodů, který měli závodníci přidělený v době konání závodu. Modifikací předchozích vzorečků vzniknou následující dva vzorečky

$$Bz = Rp - Rs * \frac{\check{C}z - \check{C}p}{\check{C}s}$$

$$Bz = Rp * \left( 2 - \frac{\check{C}z}{\check{C}p} \right)$$

Pro první vzoreček, tedy i pro všechny z něj odvozené vzorečky jsem předpokládala, že čas závodníků v kategorii závodu má normální rozdělení. Tento předpoklad není úplně pravdivý. Často doběhne většina závodníků v relativně podobném čase a několik zbloudilých závodníků s dlouhými ztrátami. Při normálním rozdělení časů závodníků by měl rozdíl průměrného času a času prvních závodníků být srovnatelný s rozdílem průměrného času a času posledních závodníků. Poslední závodníci dobíhají s daleko většími ztrátami na průměrný čas a celkově tento čas snižují výrazněji, než jej první závodníci zvyšují. Tento problém by mohlo vyřešit zlogaritmování časů závodníků. Vzorečky používající zlogaritmování času jsou tyto

$$Bz = 1000 - 200 * \frac{\log(\check{C}z) - \check{C}p_{\log}}{\check{C}s_{\log}}$$

$$Bz = 1000 * \left( 2 - \frac{\log(\check{C}z)}{\check{C}p_{\log}} \right)$$

Podle všech navrhovaných vzorečků bych měla občas některým závodníkům přidělit záporný počet bodů, ale to bych silně znevýhodňovala několik málo závodníků, kterým se nějaký závod výrazně nepovedl. Rozhodla jsem se všem těmto závodníkům přidělit nízký kladný počet bodů, konkrétně 20. K vysokému rankingu jim nepomůžou, ale ani nebudou příliš škodit, jsou to takové body za snahu a úspěšné dobehnutí závodu.

## 7.5 Současný český ranking

Vzoreček, na který nesmíme zapomenout, je ten, podle kterého se počítá současný český ranking pro hlavní kategorie závodů, tedy kategorie H21 a D21. Ranking je počítán zvlášť pro muže a ženy, není rozdělen do žádných dalších podkategorií. Závodníci různých věkových skupin jsou porovnáváni dohromady a pokud se chtějí umístit v rankingu jsou nuceni běhat v nejvýkonnější věkové kategorii závodů.

Ranking je počítán jako součet bodů 10 nejlépe ohodnocených závodů za poslední rok. Pokud závodník běžel za dané období méně než 10 závodů, ranking se spočte jako součet bodů všech jeho závodů. Následně je znormován tak, aby první mužský a ženský závodník obdržel 10 000 bodů.

Vzoreček pro výpočet bodů přidělených závodníkovi za závod je následující

$$B_z = \left( 2 - \frac{\check{C}_z}{\check{C}_{p_3}} \right) * Rp_5 * \left( 1 - K_s * \frac{U - 1}{N - 1} \right) * Kk ,$$

kde  $B_z$  jsou body přidělené závodníkovi za závod  
 $\check{C}_z$  je čas závodníka  
 $\check{C}_{p_3}$  je průměrný čas 3 nejrychlejších závodníků zařazených do rankingu  
 $Rp_5$  je průměrný ranking 5 závodníků s nejvyšším rankingem  
 $K_s$  je faktor ovlivňující výpočet bodů  
 $U$  je umístění závodníka  
 $N$  je počet závodníků v kategorii závodu  
 $Kk$  je koeficient závodu

Faktor ovlivňující výpočet bodů má pro intervalový start hodnotu 0, pro handicapový start hodnotu 0,08 a hromadný start hodnotu 0,15.

Koeficient závodu je v rozmezí od 1,00 do 1,10 popisuje důležitost závodu. Oblastní závody mívají koeficienty 1,00, žebříčky Čech a Moravy 1,03, celorepublikové žebříčky 1,06 a například finále mistrovství České republiky na krátké a klasické trati má koeficient závodu 1,10.



## 7.6 Některé další rankiny

Mezinárodní federace orientačního běhu počítá celosvětový ranking pro kvalitní závodníky hlavních kategorií. Je velmi prestižní běžet v závodě zařazeném do tohoto rankingu. Ranking závodníka je spočítán jako součet čtyř nejvyšších bodů získaných za závody. Body přidělené závodníkům za jednotlivé závody jsou počítány podle vzorečku

$$Bz = \left( Rp + Rs * \frac{\check{C}p - \check{C}z}{\check{C}s} \right) * Kk$$

Časové a rankingové průměry a směrodatné odchylky jsou počítány z časů a rankingů všech závodníků, kterým byl přidělen ranking. Pro započítání výsledků závodu je třeba, aby se ho zúčastnilo alespoň 6 závodníků s přiděleným rankingem.

Koeficient závodu  $Kk$  je obvykle stanoven na 1,00. Pro finále mistrovství světa v orientačním běhu je tento koeficient stanoven na 1,10, pro kvalifikační závody mistrovství světa, pro finále světových her a finále A soutěže World Cup je koeficient závodu 1,05.

V závodech hodnocených tímto rankingem je výrazně vyšší kvalita a nasazení závodníků ve srovnání s veteránskými kategoriemi oblastních závodů v České republice. Přesto je možné se tímto vzorcem nechat inspirovat.

Švédský ranking používá opačný systém bodování závodníků, nižší ranking znamená lepší ohodnocení závodníka. Ranking závodníka je spočítán jako průměr šesti nejlepších, tedy nejnižších, bodů ze závodů, které se konaly během posledního roku. Ranking je aktualizován třikrát za rok, konkrétně 1. března, 1. května a 1. srpna.

Body přidělené závodníkovi za závod jsou počítány podle následujícího vzorečku.

$$Bz = \check{C}z * \frac{M\check{z} + Rp + Kk}{\check{C}p} - M\check{z}$$

- kde  $Bz$  jsou body přidělené závodníkovi za závod  
 $\check{C}z$  je čas závodníka uvedený v minutách a setinách minut  
 $\check{C}p$  je průměrný čas 3 nejrychlejších závodníků zařazených do rankingu, tento čas může být maximálně 1,1 násobek času vítěze  
 $Rp$  je průměrný ranking 3 závodníků s nejlepším, v tomto případě nejnižším, rankingem, příkladem dobrého rankingu je ohodnocení 0,38 nebo 1,11. Výsledky závodu, ve kterém startovali méně než tři závodníci s rankingovým ohodnocením, nebudou započítány do rankingu.  
 $M\check{z}$  je korekční koeficient nabývající pro ženskou kategorii závodu hodnoty 60 a pro mužskou kategorii závodu hodnotu 75  
 $Kk$  je koeficient závodu. Tento koeficient je zvolen jako -2 pro švédská mistrovství a elitní závody, naopak pro nevýznamné závody oblastního charakteru je zvolen jako 2, v ostatních případech je zvolen jako 0.

Švédský způsob výpočtu rankingu je značně odlišný od českého způsobu. Pro lepší přijetí veteránského rankingu orientačními běžci by bylo vhodné, aby se navržený vzoreček pro výpočet rankingu alespoň v základních rysech podobal současnému. Švédský vzoreček uvádím jen pro ilustraci možné rozmanitosti způsobů výpočtu rankingu.

# Kapitola 8

## Vyhodnocení

### 8.1 Hodnocená data

Hodnocená data pocházejí z výsledků všech závodů, které se konaly v roce 2008, a také z hodnot získaných při výpočtu bodů přidělených závodníkům za závod a při výpočtu rankingu. Pro závodníka na závodě zpracovávám jeho čas, umístění, body přidělené za závod, ranking, který měl závodník přidělený v době závodu, a pořadí závodníka v seřazení podle času nebo rankingu. Pro kategorii závodu je navíc spočítán průměrný čas, směrodatná odchylka času, průměrný ranking, směrodatná odchylka rankingu a počet závodníků v kategorii.

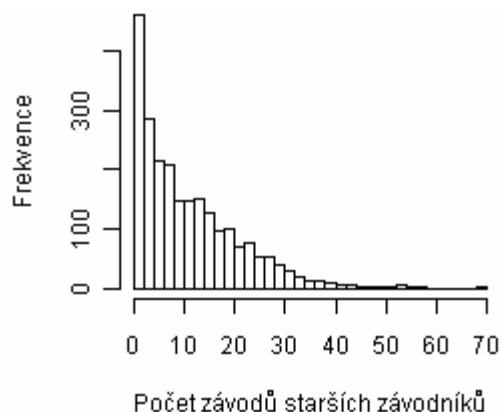
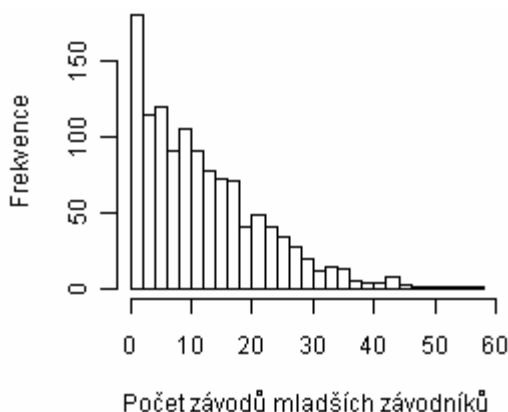
Výpočet rankingu jsem prováděla podle různých vzorečků pro závody od začátku roku 2006 do začátku roku 2009. Na začátku výpočtu jsem přidělila všem závodníkům stejný ranking, ranking pro výpočet bodů za závod jsem nastavila na 7 000 a ranking pro výsledkovou listinu jsem nastavila na nulu. S tímto rankingem jsem spočítala body získané za závody z roku 2006. Pro roky 2007 a 2008 jsem již přepočítávala ranking každý měsíc podle bodů, které závodníci získali za závody, které se konaly během posledního roku.

Vzorečky nepožívající pro výpočet bodů přidělených za závod průměrný ranking závodníků ani směrodatnou odchylku rankingu vycházejí při opakovaných výpočtech stále stejně. Vzorečky, které ranking používají, mohou vycházet různě pro různá nastavení počátečního rankingu, proto jsem tyto vzorečky počítala dvakrát. Nejprve jsem nastavila všem závodníkům stejný ranking a spustila výpočet, tím jsem získala první běh daného vzorečku. Výsledný ranking tohoto vzorečku jsem uložila jako počáteční ranking pro další výpočet a provedla jsem druhou iteraci výpočtu, tedy podle stejného vzorečku jsem znovu spočítala ranking závodníků, tak jsem získala druhý běh vzorečku.

### 8.2 Popis závodů a závodníků

Před začátkem hodnocení jednotlivých vzorečků je třeba získat přehled o datech, na kterých je výpočet rankingu založen. Tyto informace se skládají z výsledků jednotlivých závodníků na jednotlivých závodech. Výsledků závodů je mnoho, je třeba je popsat dohromady, pokud bych se měla pokusit popsat každý závod zvlášť, získala bych velké množství nepřehledných údajů, které bych stejně nakonec musela nějak zobecnit. Rozhodla jsem se proto popsat všechny závody dohromady.

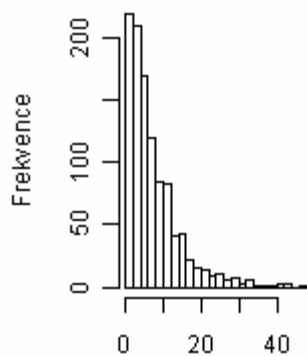
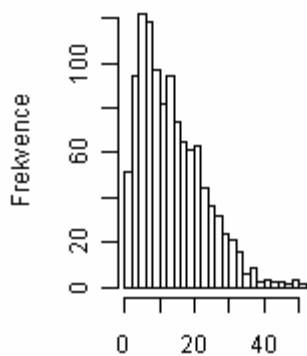
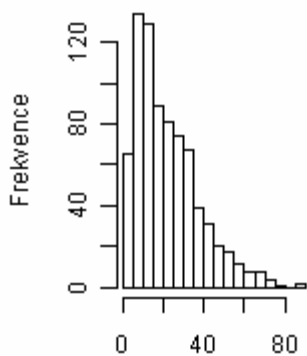
Na výsledky závodů se můžeme podívat ze dvou stran. Můžeme zkoumat výsledky jednotlivých kategorií závodu a popsat například průměrný čas a počet závodníků, kteří se závodu zúčastnili, nebo můžeme zkoumat jednotlivé závodníky a popsat počet závodů, kterých se zúčastnili.



Na těchto histogramech je zobrazen počet závodů, kterých se zúčastnili veteránští závodníci během roku 2008. Závodníky jsem rozdělila do skupin podle jejich věku. Závodníky ve věku 35 až 49 let jsem označila jako mladší závodníky a závodníky nad 50 let jsem označila jako starší.

Jen málo závodníků odběhne za rok vysoký počet závodů, navíc mladší závodníci mají častěji odběhnuto více závodů než starší. Průměr odběhnutých závodů za rok je pro mladší závodníky 12,4 a pro starší 11,4, medián je pro mladší závodníky 10 a pro starší závodníky 9 závodů za rok. Rozdíl mezi věkovými skupinami se projevuje u závodníků s nižším počtem závodů. Více než 30% závodníků neodběhne za rok více než 5 závodů u mladších závodníků a jen 3 závody u starších závodníků. Počet závodů, z kterých se závodníkům započítávají body do rankingu, je vhodné volit s ohledem na počet závodů, který závodníci odběhnou za rok.

Překvapilo mě, kolik závodů stihnou někteří závodníci za rok odběhnout. Nejvyšší počet, kterého dosáhli veteránští závodníci, je 70 závodů rok 2008, přitom je celkem 86 dní, kdy se konal nějaký závod orientačního běhu v České republice.



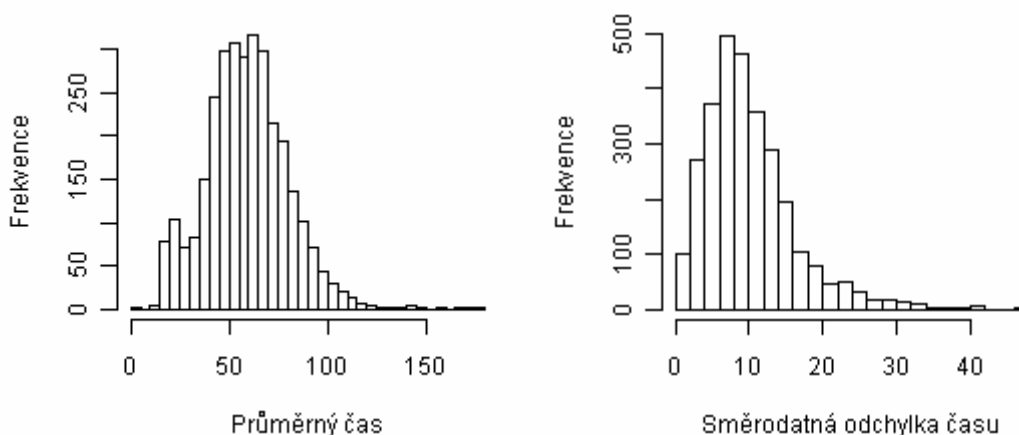
Závodníci hlavní kategorie

Závodníci silnější kategorie

Závodníci slabší kategorie

Na těchto histogramech je zobrazen počet závodníků v kategoriích jednotlivých závodů pořádaných v roce 2008. Kategorie jsem rozdělila do tří skupin. První skupinu tvoří hlavní kategorie bez věkového omezení. Druhá skupina obsahuje kategorie D35, H35 až D45, H45, které jsem označila jako silnější kategorie. Tyto kategorie jsou primárně určené pro závodníky ve věku 35 až 49 let. Poslední skupina popisuje kategorie určené pro závodníky starší 50 let, které jsem označila jako slabší kategorie.

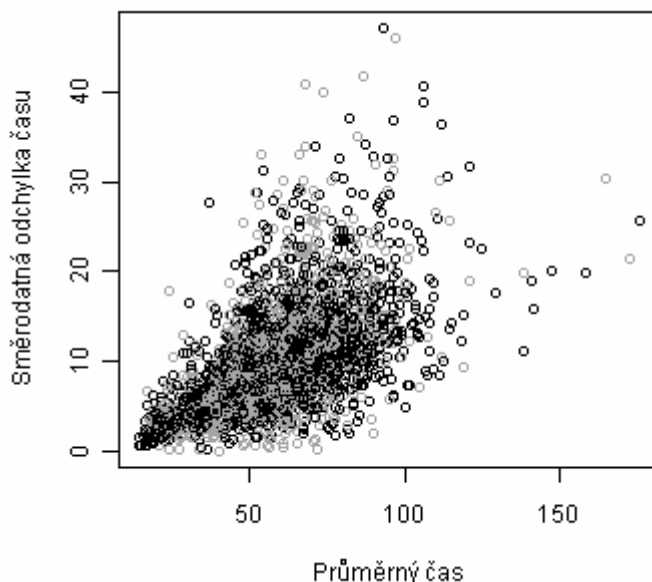
Je vidět, že kategorie starších závodníků bývají výrazně méně obsazené. Průměrný počet závodníků v hlavní kategorii je 23 závodníků, pro silnější kategorie je průměr 14 závodníků a pro slabší kategorie 8 závodníků. Mediány jsou nižší vzhledem k asymetrickému rozdělení. Medián pro hlavní kategorie je 19 závodníků, pro silnější kategorie 12 závodníků a pro slabší kategorie jen 6 závodníků.



Tyto histograme znázorňující pro jednotlivé kategorie závodů průměrný čas 90% nejrychlejších závodníků kategorie a směrodatnou odchylku času těchto závodníků. Trati závodu bývají stavěné tak, aby závodníci odběhli závod přibližně za hodinu, je vidět, že se to daří, průměr průměrného času je totiž 59,1 minut, medián 58,3 minut. Průměrný čas v kategorii má přibližně normální rozdělení.

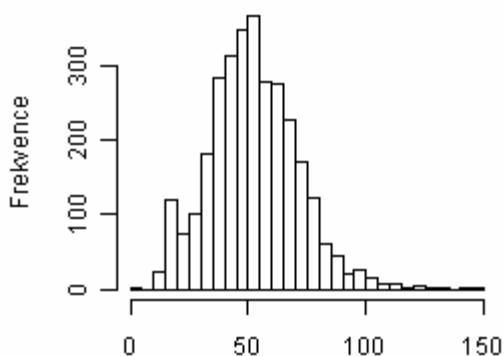
Jedním z typů orientačního běhu jsou závody ve sprintu, průměrný čas těchto závodů samozřejmě není hodina, ale přibližně 20 minut, jak je vidět z grafu. Další typ závodů jsou závody na dlouhé trati. V grafu jsou vidět jako těžší konec rozdělení ve vyšších hodnotách průměrného času závodů, těchto závodů je poměrně málo.

Směrodatná odchylka udává o kolik se průměrně liší časy závodníků od průměrného času kategorie. Medián směrodatné odchylky je 8,5 minuty, to znamená, že přibližně v polovině kategorií závodů doběhly přibližně dvě třetiny závodníků v čase, lišícím se od průměrného času kategorie nejvýše o 8,5 minuty, pokud považujeme čas závodníků v kategorii za normální. Průměr směrodatné odchylky je 9,5 minuty.

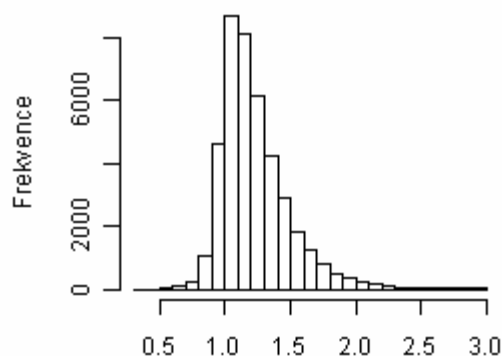


Na tomto grafu je zobrazena závislost průměrného času 90% nejrychlejších závodníků kategorie závodu a směrodatné odchylky jejich časů. Korelační koeficient těchto veličin je 0,64. Pokud bychom se omezili na kategorie s více závodníky, byl by korelační koeficient vyšší, například pro kategorie s alespoň 12 závodníky je korelační koeficient již 0,73.

Tmavě jsou vykresleny mužské kategorie a světle ženské. Je vidět, že vztah průměrného času závodníků kategorie a směrodatné odchylky časů kategorie je podobný pro obě pohlaví. Mohu tedy používat stejný vzoreček pro mužské i ženské kategorie.



Průměrný čas 4 nejrychlejších závodníků



Čas závodníka / Průměrný čas 4 nejrychlejších

První histogram zobrazuje průměrný čas čtyř nejrychlejších závodníků v kategorii závodu. Průměrný čas prvních čtyř závodníků kategorie má téměř normální rozdělení stejně jako na průměrný čas většiny závodníků. I zde jsou pozorovatelné výsledky sprintů a závodů na dlouhých tratích. Průměrný čas čtyř nejlepších závodníků je 52,3 minuty, medián je 51,2 minuty. Tyto časy jsou nižší než průměrné časy většiny závodníků, jak se dalo předpokládat.

Druhý histogram zobrazuje podíl času závodníků a průměrného času čtyř nejlepších závodníků kategorie. Několik málo závodníků má lepší čas než tento průměr, jsou to právě ti, kteří skončili s předstihem první, druhí nebo vyjíměčně i třetí. Ostatní závodníci mají čas odstupňovaný podle ztráty na vítězné závodníky. Průměr je 1,23 násobkem času nejlepších čtyř závodníků, medián je 1,17 násobkem.

Čas horší než 2 násobek průměrného času vítězů má jen přibližně 2% závodníků. Podle variant vzorečků, používajících podíl času závodníka a průměrného času prvních 4 závodníků, těmto závodníkům nebudou přiděleny téměř žádné body za závod. Je jich poměrně málo, proto myslím, že nemá cenu se tímto problémem zabývat.

### 8.3 Modifikovaný Spearmanův korelační koeficient

Vhodnost jednotlivých vzorečků pro výpočet rankingu jsem porovnávala pomocí modifikovaného Spearmanova korelačního koeficientu. Pokud je pořadí závodníků podle rankingu podobné pořadí ve kterém závodníci doběhli, pak vzoreček přiřazuje ranking odpovídající výkonnosti závodníků. Vyšší koeficient znamená lepší shodu.

V následujících tabulkách jsou zapsány hodnoty koeficientu pro ranking spočtený různými způsoby, kde body přidělené za závod byly vypočítány podle různých formulí, průměr a směrodatná odchylka času a rankingu byla počítána z časů a rankingů různého počtu nejlepších závodníků s přiděleným rankingem. Korelační koeficienty jednotlivých rankingů byly počítány jen ze závodníků s přiřazeným patřičným rankingem.

Následující tabulky jsou rozděleny do dvou částí. V horní polovině je vypočten modifikovaný Spearmanův korelační koeficient času závodníků a rankingu používaného při výpočtu bodů za závod. Tento ranking by měl odpovídat výkonnosti jednotlivých závodníků. V druhé polovině je vypočten modifikovaný Spearmanův korelační koeficient času závodníků a rankingu používaného pro sestavování výsledkové listiny závodníků. Tento ranking znevýhodňuje závodníky s nízkým počtem odběhnutých závodů a proto má nižší korelační koeficient. Pokud by některému závodníkovi vadilo toto znevýhodnění, stačí mu běhat více závodů a znevýhodnění zmizí.

Tabulka modifikovaných Spearmanových korelačních koeficientů pro různé vzorečky.

Počet nejlepších závodníků, z jejichž výsledků je počítán průměr a rozptyl času a rankingu	90% závodníků	75% závodníků	50% závodníků
Ranking počítaný jako průměr bodů za nejlepších 5 závodů, minimálně 3 odběhnuté závody			
$B_z = 1000 - 200 * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	0,7872	0,7851	0,7811
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$   1. běh	0,8267	0,8255	0,8247
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$   2. běh	0,8164	0,8204	0,8287
$B_z = 1000 - 200 * \frac{\log(\check{C}_z) - \check{C}_{p_{\log}}}{\check{C}_{s_{\log}}}$	0,7869	0,7849	0,7807
$B_z = 1000 * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	0,7541	0,7627	0,7731
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$   1. běh	0,8125	0,8148	0,8200
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$   2. běh	0,8130	0,8160	0,8137
$B_z = 1000 * \left( 2 - \frac{\log(\check{C}_z)}{\check{C}_{p_{\log}}} \right)$	0,7651	0,7705	0,7785
Ranking počítaný jako součet bodů za nejlepších 5 závodů, minimálně 1 odběhnutý závod			
$B_z = 1000 - 200 * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	0,7626	0,7631	0,7638
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$   1. běh	0,7946	0,7933	0,7927
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$   2. běh	0,7840	0,7876	0,7950
$B_z = 1000 - 200 * \frac{\log(\check{C}_z) - \check{C}_{p_{\log}}}{\check{C}_{s_{\log}}}$	0,7620	0,7624	0,7624
$B_z = 1000 * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	0,7305	0,7389	0,7490
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$   1. běh	0,7839	0,7864	0,7914
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$   2. běh	0,7837	0,7850	0,7844
$B_z = 1000 * \left( 2 - \frac{\log(\check{C}_z)}{\check{C}_{p_{\log}}} \right)$	0,7357	0,7406	0,7480

Tabulka modifikovaných Spearmanových korelačních koeficientů pro různé vzorečky.

Počet nejlepších závodníků, z jejichž výsledků je počítán průměr a rozptyl času a rankingu		3 závodníci	4 závodníci	5 závodníků
		Ranking počítaný jako průměr bodů za nejlepších 5 závodů, minimálně 3 odběhnuté závody		
$B_z = 1000 - 200 * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$		0,7281	0,7285	0,7284
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	1. běh	0,7672	0,8030	0,8139
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	2. běh	0,7862	0,8140	0,8260
$B_z = 1000 - 200 * \frac{\log(\check{C}_z) - \check{C}_{p_{\log}}}{\check{C}_{s_{\log}}}$		0,7257	0,7524	0,7271
$B_z = 1000 * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$		0,7567	0,7486	0,7422
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	1. běh	0,8044	0,8022	0,7993
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	2. běh	0,7935	0,7953	0,7945
$B_z = 1000 * \left( 2 - \frac{\log(\check{C}_z)}{\check{C}_{p_{\log}}} \right)$		0,7594	0,7524	0,7463
		Ranking počítaný jako součet bodů za nejlepších 5 závodů, minimálně 1 odběhnutý závod		
$B_z = 1000 - 200 * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$		0,7210	0,7207	0,7202
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	1. běh	0,7475	0,7767	0,7855
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	2. běh	0,7648	0,7861	0,7956
$B_z = 1000 - 200 * \frac{\log(\check{C}_z) - \check{C}_{p_{\log}}}{\check{C}_{s_{\log}}}$		0,7186	0,7185	0,7182
$B_z = 1000 * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$		0,7357	0,7287	0,7224
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	1. běh	0,7792	0,7770	0,7743
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	2. běh	0,7694	0,7707	0,7695
$B_z = 1000 * \left( 2 - \frac{\log(\check{C}_z)}{\check{C}_{p_{\log}}} \right)$		0,7310	0,7250	0,7196

Tabulka modifikovaných Spearmanových korelačních koeficientů pro různé vzorečky.

Počet nejlepších závodníků, z jejichž výsledků je počítán průměr a rozptyl času a rankingu	90% závodníků	75% závodníků	50% závodníků
Ranking počítaný jako průměr bodů za nejlepších 7 závodů, minimálně 4 odběhnuté závody			
$B_z = 1000 - 200 * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	0,7860	0,7840	0,7807
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$   1. běh	0,8260	0,8233	0,8234
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$   2. běh	0,8219	0,8208	0,8188
$B_z = 1000 - 200 * \frac{\log(\check{C}_z) - \check{C}_{p_{\log}}}{\check{C}_{s_{\log}}}$	0,7858	0,7840	0,7801
$B_z = 1000 * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	0,7581	0,7651	0,7768
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$   1. běh	0,8173	0,8206	0,8242
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$   2. běh	0,8180	0,8199	0,8199
$B_z = 1000 * \left( 2 - \frac{\log(\check{C}_z)}{\check{C}_{p_{\log}}} \right)$	0,7673	0,7726	0,7813
Ranking počítaný jako součet bodů za nejlepších 7 závodů, minimálně 1 odběhnutý závod			
$B_z = 1000 - 200 * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	0,7513	0,7536	0,7573
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$   1. běh	0,7778	0,7761	0,7766
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$   2. běh	0,7735	0,7728	0,7712
$B_z = 1000 - 200 * \frac{\log(\check{C}_z) - \check{C}_{p_{\log}}}{\check{C}_{s_{\log}}}$	0,7508	0,7525	0,7555
$B_z = 1000 * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	0,7246	0,7312	0,7422
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$   1. běh	0,7748	0,7700	0,7824
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$   2. běh	0,7750	0,7766	0,7767
$B_z = 1000 * \left( 2 - \frac{\log(\check{C}_z)}{\check{C}_{p_{\log}}} \right)$	0,7252	0,7299	0,7374



Tabulka modifikovaných Spearmanových korelačních koeficientů pro různé vzorečky.

Počet nejlepších závodníků, z jejichž výsledků je počítán průměr a rozptyl času a rankingu		3 závodníci	4 závodníci	5 závodníků
		Ranking počítaný jako průměr bodů za nejlepších 7 závodů, minimálně 4 odběhnuté závody		
$B_z = 1000 - 200 * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$		0,7272	0,7282	0,7270
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	1. běh	0,7463	0,7947	0,8088
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	2. běh	0,7721	0,7981	0,8075
$B_z = 1000 - 200 * \frac{\log(\check{C}_z) - \check{C}_{p_{\log}}}{\check{C}_{s_{\log}}}$		0,7241	0,7254	0,7259
$B_z = 1000 * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$		0,7639	0,7570	0,7509
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	1. běh	0,8074	0,8062	0,8034
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	2. běh	0,7977	0,7984	0,7977
$B_z = 1000 * \left( 2 - \frac{\log(\check{C}_z)}{\check{C}_{p_{\log}}} \right)$		0,7660	0,7602	0,7548
		Ranking počítaný jako součet bodů za nejlepších 7 závodů, minimálně 1 odběhnutý závod		
$B_z = 1000 - 200 * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$		0,7188	0,7188	0,7166
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	1. běh	0,7202	0,7578	0,7680
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	2. běh	0,7427	0,7594	0,7644
$B_z = 1000 - 200 * \frac{\log(\check{C}_z) - \check{C}_{p_{\log}}}{\check{C}_{s_{\log}}}$		0,7159	0,7158	0,7145
$B_z = 1000 * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$		0,7336	0,7273	0,7216
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	1. běh	0,7717	0,7700	0,7673
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	2. běh	0,7637	0,7637	0,7626
$B_z = 1000 * \left( 2 - \frac{\log(\check{C}_z)}{\check{C}_{p_{\log}}} \right)$		0,7253	0,7201	0,7154

Tabulka modifikovaných Spearmanových korelačních koeficientů pro různé vzorečky.

Počet nejlepších závodníků, z jejichž výsledků je počítán průměr a rozptyl času a rankingu		90% závodníků	75% závodníků	50% závodníků
		Ranking počítaný jako průměr bodů za nejlepších 10 závodů, minimálně 5 odběhnutých závody		
$Bz = Rp - Rs * \frac{\check{C}z - \check{C}p}{\check{C}s}$	1. běh	0,8240	0,8251	0,8231
$Bz = Rp * \left(2 - \frac{\check{C}z}{\check{C}p}\right)$	1. běh	0,8241	0,8264	0,8288
		Ranking počítaný jako součet bodů za nejlepších 10 závodů, minimálně 1 odběhnutý závod		
$Bz = Rp - Rs * \frac{\check{C}z - \check{C}p}{\check{C}s}$	1. běh	0,7463	0,7477	0,7481
$Bz = Rp * \left(2 - \frac{\check{C}z}{\check{C}p}\right)$	1. běh	0,7569	0,7598	0,7639

Tabulka modifikovaných Spearmanových korelačních koeficientů pro různé vzorečky.

Počet nejlepších závodníků, z jejichž výsledků je počítán průměr a rozptyl času a rankingu		3 závodníci	4 závodníci	5 závodníků
		Ranking počítaný jako průměr bodů za nejlepších 10 závodů, minimálně 5 odběhnutých závodů		
$Bz = Rp - Rs * \frac{\check{C}z - \check{C}p}{\check{C}s}$	1. běh	0,7647	0,7936	0,8026
$Bz = Rp * \left(2 - \frac{\check{C}z}{\check{C}p}\right)$	1. běh	0,8121	0,8119	0,8026
		Ranking počítaný jako součet bodů za nejlepších 10 závodů, minimálně 1 odběhnutý závod		
$Bz = Rp - Rs * \frac{\check{C}z - \check{C}p}{\check{C}s}$	1. běh	0,7173	0,7339	0,7389
$Bz = Rp * \left(2 - \frac{\check{C}z}{\check{C}p}\right)$	1. běh	0,7570	0,75575	0,7540

Tabulka modifikovaných Spearmanových korelačních koeficientů pro vzoreček počítaný podle formule českého vzorečku pro všechny kategorie závodů.

Ranking počítaný jako součet bodů za nejlepších 10 závodů	Průměrný čas 3 závodníci	$Bz = Rp * \left(2 - \frac{\check{C}z}{\check{C}p}\right)$	1. běh	0,7184
	Průměrný ranking 5 závodníků	$Bz = Rp * \left(2 - \frac{\check{C}z}{\check{C}p}\right)$	2. běh	0,7229

Tabulka modifikovaných Spearmanových korelačních koeficientů pro různé vzorečky, do výpočtu korelačního koeficientu byly zahrnuty jen výsledky závodníků, kteří měli v době konání závodu započteny do rankingu body za minimálně 5 nejlepších závodů

Ranking počítaný jako průměr z bodů za závody ( minimální počet závodů )		Ranking 5 (3)	Ranking 7 (4)	Ranking 10(5)
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	3 závodníci	0,7689	0,7459	0,7647
	4 závodníci	0,8042	0,7954	0,7936
	5 závodníků	0,8147	0,8092	0,8026
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	90% závodníků	0,8267	0,8257	0,8240
	75% závodníků	0,8248	0,8230	0,8251
	50% závodníků	0,8240	0,8230	0,8231
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	3 závodníci	0,8022	0,8064	0,8121
	4 závodníci	0,8001	0,8052	0,8119
	5 závodníků	0,7971	0,8022	0,8026
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	90% závodníků	0,8109	0,8165	0,8241
	75% závodníků	0,8136	0,8199	0,8264
	50% závodníků	0,8185	0,8236	0,8288

Tabulka modifikovaných Spearmanových korelačních koeficientů pro různé vzorečky, při výpočtu průměru a rozptylu času kategorie se používají časy všech závodníků.

4 závodníci		Ranking jako průměr 7 (4)	Ranking jako součet 7 (1)
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	1. běh	0,7894	0,7540
$B_z = R_p - R_s * \frac{\check{C}_z - \check{C}_p}{\check{C}_s}$	2. běh	0,7882	0,7529
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	1. běh	0,8062	0,7712
$B_z = R_p * \left( 2 - \frac{\check{C}_z}{\check{C}_p} \right)$	2. běh	0,7981	0,7646

Nyní je třeba vyvodit nějaké závěry z vypočtených modifikovaných Spearmanových korelačních koeficientů. Je možné určit, která formule vzorečku se pro výpočet bodů přidělených za závod hodí lépe, jaký je optimální počet závodníků, z jejichž časů a rankingů se budou počítat průměry a směrodatné odchylky času a rankingu kategorie závodu, a z kolika nejvyšších bodů získaných za závody by bylo vhodné počítat ranking závodníka. Rozdíly mezi jednotlivými způsoby výpočtu nejsou příliš výrazné, žádný vzoreček není výrazně nejhorší.

Nejvýraznější rozdíly jsou mezi různými formulemi vzorečku pro výpočet bodů přidělených závodníkovi za závod. Počítala jsem podle dvou základních vzorečků upravených do několika variant. První varianta nezohledňuje různou kvalitu závodníků na závodě. Korelační koeficient těchto vzorečků vychází nižší než u propracovanějších variant. Druhá varianta vzorečku používá logaritmus času závodníků. Korelační koeficienty těchto vzorečků vycházejí podobně jako u předchozí varianty. Formule pro výpočet bodů za závod s použitím zlogaritmování času závodníků se proto zdá být zbytečně složitá.

Poslední varianta vzorečku, se kterou jsem pracovala, je používá průměrný ranking nejlepších závodníků kategorie, případně také směrodatnou odchylku jejich rankingů. Korelační koeficienty vzorečků zohledňujících kvalitu závodníků vycházejí vyšší než u předchozích úprav. Nejlepší vzorečky z navržených možností se zdají být ty, které zohledňují kvalitu závodníků na závodech. Pro lepší přehlednost jsem v některých tabulkách uvedla jen korelační koeficienty těchto vzorečků.

Základní typy vzorečků jsou dva. První typ vzorečku používá směrodatnou odchylku, druhý typ nikoliv. Jednoznačně určit, který ze vzorečků je vhodnější není úplně snadné. Pro varianty s pevným počtem závodníků se zdají být lepší vzorečky, které používají pouze průměr času a rankingu, zřejmě proto, že směrodatná odchylka počítaná ze 3 nebo 4 hodnot nemá příliš vysokou vypovídající schopnost. Pro varianty s relativním počtem závodníků, kde se počítají průměry a směrodatné odchylky z většiny závodníků, se zdají naopak být lepší vzorečky využívající směrodatné odchylky.

Další variabilní část vzorečků je počet závodníků, z jejichž časů a rankingů se počítá průměrný čas a ranking kategorie závodu, případně i jejich směrodatné odchylky. Pro vzorečky využívající směrodatné odchylky se zdá být vhodné používat co největší počet závodníků, naopak, pro vzorečky využívající pouze průměry mají větší korelační koeficienty varianty s nižším počtem závodníků. Tyto pozorování jsou shodné pro relativně i absolutně zvolený počet nejlepších závodníků.

Poslední volitelnou částí vzorečku, kterou jsem používala, je počet nejlepších závodů, jejichž body jsou započítány do rankingu. Ranking jsem počítala z 5, 7 nebo 10 nejlepších bodů, které závodník získal za poslední rok, přičemž k udělení jednotlivých rankingů, počítaných jako průměr nejlepších závodů, je třeba, aby závodník odběhl minimálně 3, 4 nebo 5 závodů. Pro porovnání vhodnosti jednotlivých variant jsem vytvořila samostatnou tabulku. V této tabulce jsem do výpočtu modifikovaných Spearmanových korelačních koeficientů zařadila jen ty závodníky, kteří měli v době konání závodu přiděleny všechny tyto rankingy, to jsou ti závodníci, kteří měli v době konání jednotlivých závodů odběhnutých alespoň 5 závodů během posledního roku. Pokud bych porovnávala korelační koeficienty vypočtené na základě všech závodníků, kteří měli v době konání přidělený ranking, získala bych neporovnatelné hodnoty.

Vzoreček využívající k výpočtu bodů za závod směrodatnou odchylku času a rankingu závodníků vychází mírně lépe při výpočtu rankingu z bodů z nižšího počtu nejlepších závodů. Pro vzoreček využívající pouze průměr času a rankingu závodníků vycházejí korelační koeficienty naopak vyšší při použití vyššího počtu nejlepších závodů.

Nejvyšší modifikovaný Spearmanův korelační koeficient ze zkoumaných možností vychází pro vzoreček, který používá při výpočtu bodů za závod průměr a směrodatnou odchylku času a rankingu 90% nejlepších závodníků a který počítá ranking za použití 5 nejvyšších bodů přidělených závodníkovi za závody během posledního roku. Dobrým návrhem na vzoreček pro výpočet veteránského rankingu orientačních běžců se zdá být volba tohoto vzorečku a jeho parametrů.

Pokud by jsme chtěli zvolit jednodušší vzoreček nebo vzoreček podobnější stávajícímu českému vzorečku pro výpočet rankingu, ať už použitím pevně zvoleného počtu nejlepších závodníků pro výpočet průměrného času a rankingu kategorie závodu nebo podobnou formulí pro výpočet bodů přidělených za závod, pak vychází nejvyšší korelační koeficient pro vzoreček používající výsledky tří nejlepších závodníků pro výpočet průměrného času a rankingu kategorie závodu a počítající ranking z bodů získaných za 10 nejlepších závodů, přičemž ranking pro výpočet bodů za závod je udělen pouze závodníkům, kteří mají odběhnuto více než 5 závodů za poslední rok.

Je třeba si důkladně rozmyslet, zda budeme trvat na tak vysokém počtu odběhnutých závodů. Musíme vzít v úvahu, že medián počtu odběhnutých závodů za rok je právě požadovaných deset závodů a také, že více než 30% závodníků starších 50 let neodběhne za rok více než tři závody. Mohlo by se snadno stát, že v některé starší veteránské kategorii závodů s několika závodníky, nebude mít nikdo přidělen ranking pro výpočet průměrného rankingu za závod. V takovém případě nebudou počítány body přidělené za závod, proto se výsledky tohoto závodu nezapočítají do rankingu a tudíž závodníci budou mít ještě méně závodů. Proto by bylo vhodné volit nižší počet závodů, jejichž body se započítávají do rankingu, a také nižší minimální počet závodů potřebný pro udělení rankingu pro výpočet bodů za závod.

Možná by stálo za úvahu zvolit vyšší počet závodníků, jejichž výsledky jsou použity pro výpočet průměrného času a rankingu kategorie závodu. Požitím vyššího počtu závodníků se zajistí vyšší stabilita výpočtu, nestane se tak snadno, aby například první závodník, kterému se povedlo doběhnout do cíle ve výrazně rychlejším čase než zbytku kategorie, výrazně snížil průměrný čas několika nejlepších závodníků a tím výrazně snížil body, které jsou závodníkům za tento závod přiděleny.

Pro zajímavost jsem zkusila vytvořit také vzorečky pro výpočet rankingu, kde se průměrné časy počítaly z časů všech závodníků, ne jen ze závodníků s přiděleným rankingem pro výpočet bodů jako v ostatních případech. Zvolila jsem jen variantu, kde je použit průměrný čas a ranking 4 nejlepších závodníků kategorie a ranking se počítá z bodů získaných za 7 nejlepších závodů. Takto spočítaný ranking vychází obdobně nebo hůře než při použití pouze času běžců s přiděleným rankingem. V případech, kdy skončili na prvních místech závodníci bez přiděleného rankingu, není možné dostatečně odhadnout kvalitu závodu, proto jsem v této variantě nepokračovala.

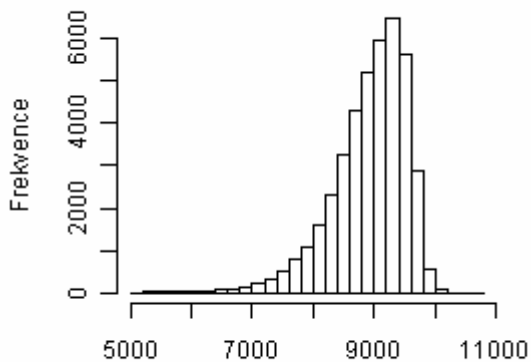
Hodnoty modifikovaného Spearmanova korelačního koeficientu jsem spočítala také podle současného českého vzorečku pro výpočet rankingu, počítala jsem ranking veteránských závodníků. Tento ranking se od ostatních liší ve dvou základních bodech. Používá pouze jeden způsob výpočtu rankingu, a to součet deseti nejvyšších bodů přidělených závodníkům za závody. Podle tohoto rankingu je počítán průměrný ranking nejlepších závodníků v závodě i výsledné pořadí závodníků. Druhý rozdíl spočívá ve volbě rozdílného počtu časově a rankingově nejlepších závodníků. Českému rankingu vycházejí nejnižší korelační koeficienty z téměř všech zkoumaných možností. Zdá se, že použití shodného počtu nejlepších závodníků pro výpočet průměru času a rankingu spolu s počítáním dvou rankingů souběžně byla dobrá volba.

## 8.4 Podrobnější popis vybraných vzorečků

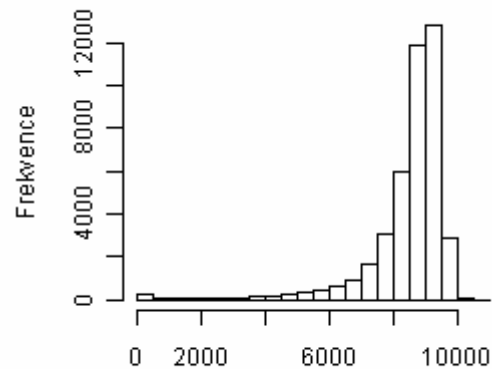
Pro lepší představu o jednotlivých vzorečcích popíši podrobněji následující dva vzorečky, které mají poměrně vysoký modifikovaný Spearmanův korelační koeficient.

Hodnotila jsem vzorečky ve kterých jsem časové a rankingové průměry a směrodatné odchylky počítala z relativně i absolutně zvoleného počtu nejlepších závodníků kategorie, vzorečky jsem vybrala tak, aby reprezentovali obě tyto skupiny.

1. vzoreček	$Bz = Rp - Rs * \frac{\check{C}z - \check{C}p}{\check{C}s}$	90% závodníků	Ranking počítaný z bodů za nejlepších 5 závodů
2. vzoreček	$Bz = Rp * \left( 2 - \frac{\check{C}z}{\check{C}p} \right)$	4 závodníci	Ranking počítaný z bodů za nejlepších 5 závodů



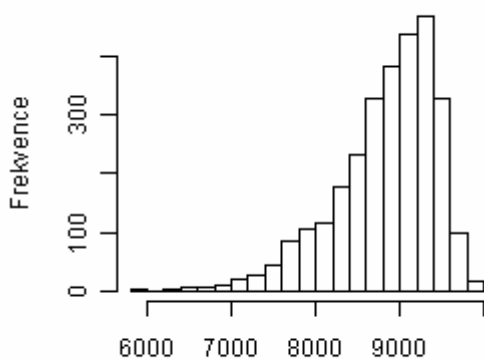
Body přidělené za závod podle 1. vzorečku



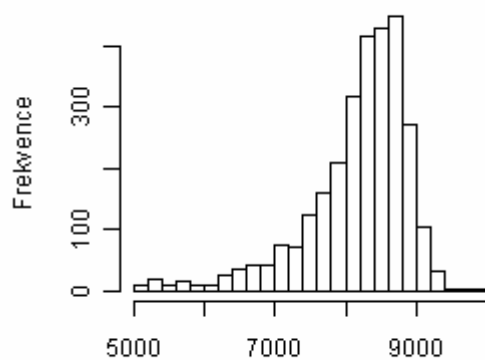
Body přidělené za závod podle 2. vzorečku

Na těchto histogramech jsou zobrazeny body přidělené závodníkům za jednotlivé závody podle prvního a druhého vzorečku. Body přidělené za závody mají asymetrické rozdělení vychýlené doprava. Vychýlení je výraznější v druhém případě, kde se časy jednotlivých závodníků porovnávají s průměrným časem čtyř nejrychlejších závodníků. Závodníci mívají větší ztrátu na první závodníky, než je jejich ztráta na průměrný čas většiny závodníků. Proto také častěji získají nižší počet bodů při výpočtu podle druhého vzorečku než při výpočtu podle prvního vzorečku.

Body přidělené závodníkům podle prvního vzorečku mají průměr 8907 bodů a směrodatnou odchylku 642 bodů. Body přidělené závodníkům podle druhého vzorečku mají průměr 8479 bodů a směrodatnou odchylku 1282 bodů.



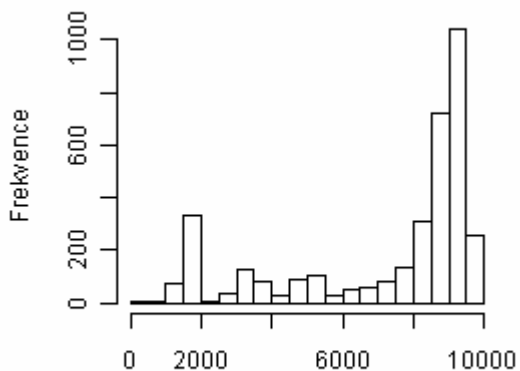
Ranking závodníků podle 1. vzorečku



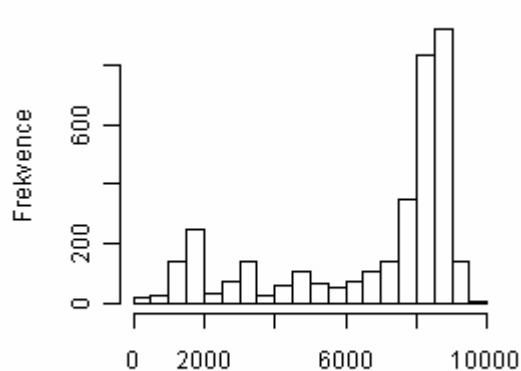
Ranking závodníků podle 2. vzorečku

Na těchto histogramech je zobrazen ranking jednotlivých závodníků spočítaný jako průměr bodů získaných z pěti nejlepších závodů. Ranking byl přidělen pouze závodníkům, kteří měli alespoň 3 odběhnuté závody. Rozdělení rankingů je vychýlené doprava. To znamená, že je několik málo závodníků s velmi vysokým rankingem, mnoho závodníků s vysokým rankingem, a se snižujícím se rankingem postupně ubývá počet závodníků, kterým byl tento ranking přidělen.

Průměr rankingu počítaného podle prvního vzorečku je 7222, medián je 8775 a směrodatná odchylka 3454. Ranking spočítaný podle druhého vzorečku je celkově nižší, průměr je 6645, medián 8160 a směrodatná odchylka 3233.



Ranking závodníků podle 1. vzorečku



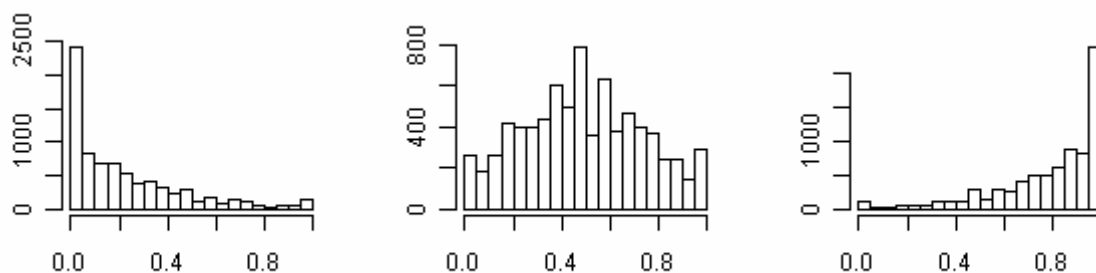
Ranking závodníků podle 2. vzorečku

Tyto histogramy zobrazují ranking jednotlivých závodníků spočtený jako součet bodů získaných z pěti nejlepších závodů jednotlivých závodníků podle obou zkoumaných vzorečků. Tyto rozdělení jsou složená rozdělení. Je krásně vidět, jak byl závodníkům s jedním, dvěma nebo třemi odběhnutými závody, přidělen nízký ranking. Dobří závodníci se čtyřmi odběhnutými závody již mohou získat podobný ranking jako horší závodníci s více závody a proto se tyto skupiny již sloučily dohromady. V rámci jednotlivých skupin závodníků je rozdělení vychýlené doprava podobně jako rozdělení rankingu počítaného pomocí průměru bodů získaných z pěti nejlepších závodů.

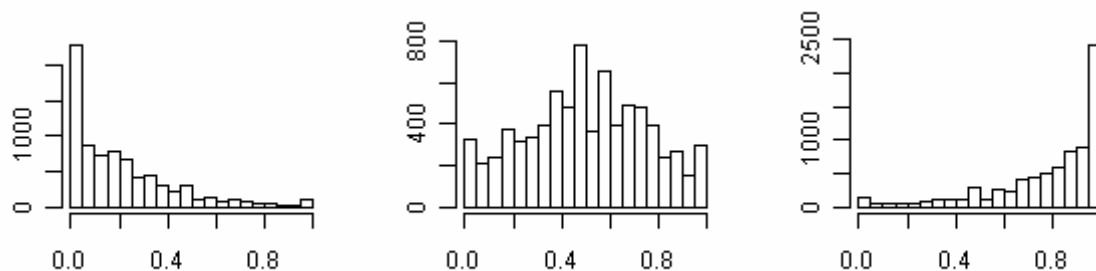
Pro všechny kategorie závodů jsem určila pořadí jednotlivých závodníků podle času závodníků a podle jejich rankingu, určeného jako průměr bodů z pěti nejlepších závodů, který měl závodník přidělený v době konání závodu. Ranking by měl vyjadřovat výkonnost jednotlivých závodníků, tedy závodníci s vysokým rankingem by se měli umísťovat na prvních místech ve výsledkové listině. Zároveň by mělo platit, že na prvních místech v jednotlivých závodech skončí závodníci s vysokým rankingem. Obdobné tvrzení by mělo platit pro závodníky umístěné uprostřed i pro závodníky posledních místech.

Rozhodla jsem se ověřit platnost všech těchto tvrzení. Pro porovnávání výsledků kategorií s různým počtem závodníků jsem použila relativní umístění závodníka. Relativní umístění je vyjádřeno předpisem  $(U-1)/(N-1)$ , kde  $U$  značí umístění závodníka podle časového nebo rankingového pořadí a  $N$  značí počet závodníků v kategorii.

Ze závodníků jsem vybrala tři skupiny nejprve podle relativního rankingového umístění a poté podle relativního časového umístění. Skupiny jsem vytvářela podle 20%, 40%, 60% a 80% kvantilů příslušného relativního umístění. Závodníky jejichž relativní umístění bylo nižší než 20% kvantil jsem označila jako do dobré, závodníky jejichž relativní umístění bylo mezi 40% a 60% kvantilem jsem označila jako prostřední a závodníky jejichž relativní umístění bylo horší než 80% kvantil jsem označila jako závodníky se špatným relativním umístěním.

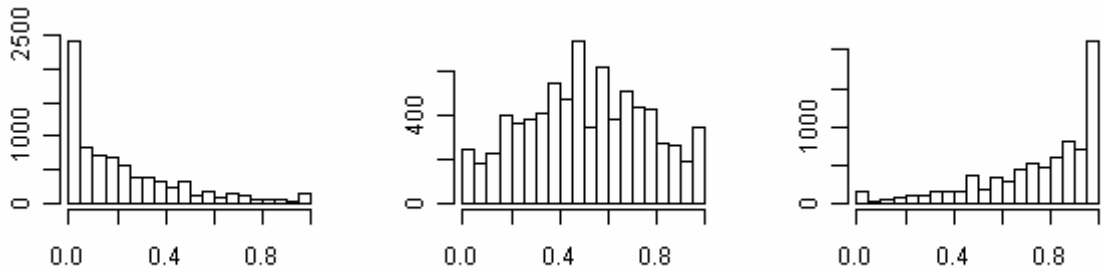


Vybrala jsem závodníky s dobrým, prostředním a špatným relativním umístěním podle rankingu spočítaného prvním vzorečkem. Pro každou skupinu jsem nakreslila histogram jejich relativního časového umístění.

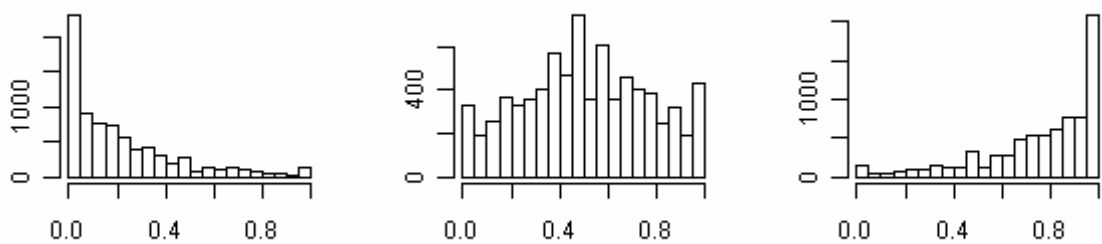


Vybrala jsem závodníky s dobrým, prostředním a špatným relativním časovým umístěním. Pro každou skupinu jsem nakreslila histogram jejich relativního umístění podle rankingu spočítaného prvním vzorečkem.





Vybrala jsem závodníky s dobrým, prostředním a špatným relativním umístěním podle rankingu spočítaného druhým vzorečkem. Pro každou skupinu jsem nakreslila histogram jejich relativního časového umístění.

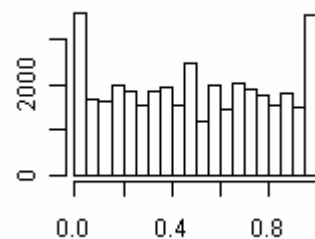


Vybrala jsem závodníky s dobrým, prostředním a špatným relativním časovým umístěním. Pro každou skupinu jsem nakreslila histogram jejich relativního umístění podle rankingu spočítaného druhým vzorečkem.

Z uvedených histogramů je pěkně vidět, že závodníci dobře umístění podle jedné varianty pořadí se umísťují na dobrých místech i podle druhé varianty pořadí. Stejně tak závodníci se špatným umístěním podle jedné varianty pořadí se umísťují na koncových místech podle druhé varianty pořadí.

Pro závodníky s prostředním umístěním by se mohl zdát výsledek méně průkazný, ale je třeba si uvědomit, že k tomu, aby vůbec nějaký závodník mohl mít relativní umístění mezi 0,51 a 0,55, je třeba aby v závodě startovalo alespoň 12 závodníků, takových závodů je poměrně málo. Tím se dá vysvětlit méně časté zastoupení některých skupin v histogramech zobrazujících relativní umístění prostředních závodníků. S přihlédnutím, k této mírné nerovnoměrnosti v zastoupení jednotlivých skupin závodníků, můžeme říci, že závodníci které měli prostřední umístění podle jednoho pořadí, velmi často skončili ve střední části závodníků podle druhého umístění.

Pro kontrolu ještě histogram zobrazující relativní pořadí všech závodníků podle času a podle rankingu spočítaného prvním i druhým vzorečkem. Histogram je ve všech těchto případech stejný, protože je vždy někdo první, někdo druhý a někdo poslední, ať už se pořadí závodníků sestavují jakkoliv. Rozdělení relativního umístění je přibližně rovnoměrné.



Z uvedených histogramů je vidět, že vypočtený ranking skutečně odpovídá odhadu výkonnosti jednotlivých závodníků.

# Kapitola 9

## Závěr

V této práci jsem navrhla a vytvořila systém pro výpočet veteránského rankingu závodníků orientačního běhu. Při navrhování vhodných vzorečků pro výpočet rankingu jsem se inspirovala vzorečky používanými v českých i zahraničních soutěžích, které jsou však pořádány převážně pro mistrovské kategorie. Veteránské kategorie jsou specifické a odlišné od mistrovských především nižším počtem závodníků a možností migrace závodníků mezi kategoriemi. S ohledem na tato kritéria jsem navrhla vzorečky vhodné pro výpočet veteránského rankingu orientačního běhu včetně možných parametrizací a provedla jejich důkladný rozbor.

Základ systému pro výpočet rankingu tvoří databáze, ve které jsou uloženy výsledky jednotlivých závodů orientačního běhu od roku 2006 do 2008, ranking závodníků spočtený podle různých vzorečků a samotné vzorečky pro výpočet rankingu. Takto vytvořená databáze se ukázala jako vhodná volba, protože umožňuje zároveň počítat ranking podle mnoha různých vzorečků.

Pro porovnání vhodnosti jednotlivých vzorečků pro výpočet rankingu jsem navrhla modifikaci Spearmanova korelačního koeficientu. Metoda spočívá v porovnávání pořadí závodníků podle času, kterého v závodě dosáhli, a pořadí podle rankingu, který jim byl přidělen podle jednotlivých vzorečků. Pokud je časové pořadí závodníků podobné jejich pořadí podle rankingu, pak přidělený ranking vyjadřuje výkonnost jednotlivých závodníků.

Pro každý vzoreček byly počítány dvě varianty rankingu současně. První se použila pro určení kvality závodníků, kteří se závodů zúčastnili, a druhá k sestavení pořadí závodníků. Ranking závodníků jsem počítala podle dvou základních vzorečků a několika jejich variant pro celkem 128 různých modifikací. Nejlépe vycházely varianty vzorečků používající průměrný ranking nejlepších závodníků kategorie a tím zohledňující kvalitu jednotlivých závodů.

Nejllepší shoda mezi pořadím podle času a podle rankingu byla dosažena pro vzoreček používající průměr a směrodatnou odchylku času a rankingu 90% nejlepších závodníků a počítající ranking za použití pěti nejvyšších bodů přidělených závodníkovi za závody během posledního roku.

Velmi dobrá shoda mezi pořadím podle času a podle rankingu vychází i pro vzoreček podobný stávajícímu českému vzorečku používající průměrný čas a ranking tří nejlepších závodníků kategorie závodu a počítající ranking závodníka z bodů získaných za 10 nejlepších závodů, přičemž ranking pro výpočet bodů za závod je udělen pouze závodníkům, kteří mají odběhnuto více než 5 závodů za poslední rok. Při volbě počtu odběhnutých závodů je třeba vzít v úvahu, že starší závodníci odběhnou za rok poměrně nízký počet závodů.

Při tvorbě bakalářské práce jsem osvědčila v praxi mnohé z naučených poznatků, rozšířila jsem své vědomosti o orientačním běhu, speciálně o problematice rankingu, zkusila si práci na rozsáhlejším projektu a tato práce mě bavila.

# Literatura

Welling L., Thomson L.: *MySQL Průvodce základy databázového systému*. CP Books, Brno, 2005.

Kosek J.: *PHP tvorba internetových aplikací*. Grada Publishing, Praha, 1999.

Forst L.: *Úvod do UNIXu*  
<http://www.ms.mff.cuni.cz/~forst/UXover.ppt>

Anděl J.: *Statistické metody*. Matfyzpress, Praha, 1993.

Anděl, J.: *Základy matematické statistiky*. Matfyzpress, Praha, 2005.

Havránek, T.: *Statistika pro biologické a lékařské vědy*. Academia, Praha, 1993.

The R Project for Statistical Computing  
<http://www.r-project.org/>

Oficiální stránka Českého svazu orientačního běhu  
<http://www.orientacnibeh.cz/>

Kalendář závodů ČSOB  
<http://klobouk.fsv.cvut.cz/~hora/kalendar/kal2001.php>

Český ranking  
[http://mujweb.cz/www/pbmob/02\\_ranking.htm](http://mujweb.cz/www/pbmob/02_ranking.htm)

British Orienteering Ranking List  
<http://ccgi.orienteeing.plus.com/rankings/html/rankhome.html>

Scottish O League Scoring System discussion  
<http://www.scottish-orienteeing.org/index/soa/page/sol-scoring-system-discussion>

About the IOF World Ranking Scheme  
[http://www.6prog.org/iof/IOF\\_Documents/FootO/wrscheme.htm](http://www.6prog.org/iof/IOF_Documents/FootO/wrscheme.htm)

Rankinglistan - Svenska Orienteringsförbundet  
<http://www.obasen.nu/ranking/meriteringsgrunder.asp>

Suomen ranki  
<http://www.ssl.fi/SSL/sslweb.nsf/sp?Open&cid=content43E99B>