



Универзитет у Новом Саду
Природно-математички факултет
Департман за математику и информатику

Процена развоја јуниорских пливача у новим и старим старосним категоријама: Статистичка анализа

Професор:
др Загорка Лозанов-Црвенковић

Студент:
Никола Попадић 344/21

Нови Сад, јул 2024. године

Садржај

1	Увод	2
2	Подаци	3
2.1	FINA поени	3
2.2	Прикупљање података	3
2.2.1	Функција <i>get_main_page_info</i>	4
2.2.2	Функција <i>scrape_best_times</i>	5
2.2.3	Функција <i>write_to_csv</i>	6
2.2.4	Сређивање базе	6
3	Статистички тестови	9
3.1	Подела типова података	9
3.2	Провера нормалне расподеле променљивих	9
3.3	Дескриптивна статистика	10
3.3.1	Дескриптивна статистика групе 1	10
3.3.2	Дескриптивна статистика групе 2	11
3.3.3	Дескриптивна статистика групе 3	12
3.3.4	Дескриптивна статистика свих група заједно	13
3.4	Једнострука линеарна регресија	13
3.5	T тест	15
3.6	ANOVA	17
4	Закључак	20

1 Увод

Пубертет је период живота ког прате психичке и физичке промене. Иако се код девојака одвија раније него код момака, већина промена које настају у овом периоду су исте за оба пола. Пораст висине и мишићне масе су само неке од видљивих промена, док део дискретнијих чине побољшање координације и моторичких вештина. Били очигледни или не, сви развоји који долазе са пубертетом значајно утичу на перформансе младих спортиста.[1]

Категорија која у пливању прати овај нагли напредак назива се *јуниорска категорија*. До 21.02.2023. Светска Пливачка Организација је под јуниорима сматрала девојке од 14 до 17 и момке од 15 до 18 година. На поменути датум је усвојен нови правилник који за оба пола изједначава јуниорски узраст на 15 до 18 година.

Идеја овог истраживања је да покаже да је већа сличност између полова код прве расподеле него код друге. Да би се то показало користе се статистички тестови који пореде или прогрес или одређене године током јуниорске категорије код следећих група. Прву групу чине момци који током јуниорске категорије имају од 15 до 18 година. То је "основна" група са којом се пореде остале 2 женске. Следећу групу чине девојке које су јуниорке по новим правилима, тј. имају 15-18 година и за њих се сматра да резултатски немају велику сличност са мушком јуниорском категоријом. Последњу групу чине девојке које су јуниорке по старим правилима, тј. имају 14-17 година и за њих се сматра да имају већу сличност са мушком јуниорском категоријом. За женске групе се користе исте девојке у различитим периодима развоја (14-17 и 15-18 година) да би се избегла пристрасност ранијег или каснијег уласка у пубертет код различитих девојака.

Сврха показивања веће сличности међу половима у првој него у другој категоризацији, а самим тим и овог истраживања је резултатска подударност полова у истој категорији. Јуниорска категорија је једина за коју се у пливању специфично одржавају такмичења од државног до светског нивоа и зато ако ће се међусобно поредити и проглашавати најуспешнији јуниори требало би да они буду на истом нивоу.

2 Подаци

2.1 FINA поени

Као и у многим мултидисциплинарним спортовима у пливању постоји метрика која служи за поређење резултата различитих дисциплина и пола. У овом истраживању та метрика представља променљиву на којој се раде тестови, стога да би се истраживање разумело круцијално је разумети и како та метрика функционише. FINA поен је бројчана вредност која се додељује сваком пливачком резултату, која фундаментално мери близину испливаног времена са светским рекордом у тој трци. Акроним долази од првог назива за интернационалну пливачку федерацију "*Fédération internationale de natation*", и сваком резултату одговора тачно једна вредност FINA поена. FINA поени се рачунају уз помоћ следеће формуле:

$$FINA = 1000 \left(\frac{B}{T} \right)^3 \quad (1)$$

где:

- B (base time) означава основно време тј. тренутни светски рекорд у трци за коју се рачунају FINA поени.
- T (time) означава испливано време за које се рачунају FINA поени.
- степеновање бројем 3 повећава осетљивост рачуна.

Предности система за рачунање FINA поена:

- **Стандардизација:** FINA поени омогућавају стандардизовано поређење перформанси пливача и пливачица у различитим дисциплинама и дужинама. То дозволјава равноправно проглашавање најбољег пливача на полној основи. Осим тога они такође омогућавају поређење полова, као и штафетних трка, што додаје још један слој стандардизације на овај већ стандардизован систем рачунања.
- **Мотивација:** Степеновање бројем 3, које се примењује у функцији за рачунање FINA поена, може дати било који реалан број у зависности од почетне вредности. Пошто испливано време (домен) не може бити негативно, тако и FINA поени (кодомен) не могу бити негативни. Из тог разлога је функција за рачунање FINA поена експоненцијално растућа. То значи да ће пливачу бити потребан мањи напредак у времену за исти напредак у поенима. Овај, пливачима генерално непознати аспект рачунања, може додатно мотивисати спортисту да настави да улаже труд.

2.2 Прикупљање података

Истраживање се ради на пливачима и пливачицама из Србије који су пливали током целе јуниорске категорије. Спортисти су ручно селектовани од стране истраживача, који је из личног знања и уз техничку проверу осигурао комплетност и конзистентност њихових наступа током јуниорске категорије. Истраживање на нивоу државе омогућава непристрасност података, због тенденције да тренери

из различитих округа одржавају сличне тренинге. Разлог за то је тенденција да тренери из исте државе завршавају исте факултете, а и често су сами пливачи који су прошли сличне школе пливања. Такође, тренери успешнијих пливача, на којима се и ради ово истраживање, проводе пуно времена на заједничким камповима, припремама и такмичењима, на којима размењују идеје и системе тренирања. Из ових разлога могуће је непристрасно поредити резултате пливача ирелевантно од њиховог пливачког клуба.

Предност код мерења пливачких времена је њихова прецизност. Пливачке трке се мере аутоматским системима које након регистровања времена проверавају лиценциране судије које паралелно обављају ручно мерење. Након тога се резултати стављају у базу и шаљу Европској Пливачкој Федерацији (LEN). Њихова целокупна база је приватна, али сајт *swimrankings.net* има приступ њој и селековано је приказује. Проблем овог сајта је то што су подаци коцентрисани и да би се приказали одређени резултати првобитно се захтева да се унесе име државе, па затим име и презиме пливача, а на крају је потребно изабрати дисциплину и дужину трке за коју се проверава прогрес. Из тог разлога је коришћена Python скрипта која аутоматски прикупља резултате.

Скрипта користи Python библиотеку *BeautifulSoup* за извлачење података из HTML фајлова (сајт *swimrankings.net*), а затим их након распоређивања складишти у CSV фајл уз помоћ истоимене Python библиотеке. Библиотеке *requests*, *re*, *os*, *time* и *random* доприносе са функцијама које упроштавају и оптимизују овај процес.

Програм на почетку садржи низ свих пливача представљених помоћу id-ова које *swimrankings.net* додељује сваком спортисти попут примарног кључа. Осим тога, дат је и почетни url, који је заједнички за сваког пливача и завршава се са "athleteId=". Дате почетне вредности омогућавају да се приступи профилу (сајту) сваког пливача тако што се на почетни url, уз помоћ for петље, редом конкатенира свачији id. Након имплементирања почетних вредности на ред долази позивање функција.

2.2.1 Функција *get_main_page_info*

Функција *get_main_page_info* са профила пливача преузима *junior_cat_old*, *junior_cat*, *name* и *gender* уз помоћу уграђених функција *BeautifulSoup* библиотеке *get* и *find*. Променљива *junior_cat_old* представља четворочлани низ, додељен само девојкама, који садржи календарске године у којима су оне имале 14, 15, 16 и 17 година, што представља женску јуниорску категорију по старим правилима. Са друге стране променљива *junior_cat* представља четворочлани низ, додељен и девојкама и момцима, који садржи календарске године у којима су они имали 15, 16, 17 и 18 година, што представља свачију јуниорску категорију по новим правилима.

```
def get_main_page_info(main_url):
    r = requests.get(main_url)
    soup = BeautifulSoup(r.content, 'html.parser')

    name_div = soup.find('div', id='name')
    img_tag = name_div.find('img')
    src = img_tag.get('src')

    cleaned_name = clean_name(name_div.text)
```

```

gender = re.search(r'gender(\d+)\.png', src).group(1)
birth_year = [int(num) for num in re.findall(r'\d+', name_div.text)]

if gender == '2':
    junior_cat_old = [birth_year[0] + 14, birth_year[0] + 15, birth_year[0]
        + 16, birth_year[0] + 17]
    junior_cat = [birth_year[0] + 15, birth_year[0] + 16, birth_year[0] +
        17, birth_year[0] + 18]
    return junior_cat_old, junior_cat, cleaned_name, gender
else:
    junior_cat = [birth_year[0] + 15, birth_year[0] + 16, birth_year[0] +
        17, birth_year[0] + 18]
    return None, junior_cat, cleaned_name, gender

```

2.2.2 Функција *scrape_best_times*

Функција *scrape_best_times* која служи за прикупљање и складиштење резултата пливача, се у зависности од пола позива 1 или 2 пута, за момке и девојке, респективно. Њој су прослеђени подаци о имену, полу и групи пливача који се чувају у почетни низ. Група пливача је помоћна променљива која олакшава тестирања у програмима за обраду података. Може имати вредност 1 за мушке јуниоре, 2 за девојке код којих се јуниорска категорија рачуна по старим правилима и 3 за девојке код којих се јуниорска категорија рачуна по новим правилима. У бази података је појам трка сачуван као 2 појма (променљиве) које представљају дужину и дисциплину трке. Функција *scrape_best_times* за сваког пливача чува резултате у 3 олимпијске трке у којима је пливач у просеку током јуниорске категорије отпливао резултате са највише FINA поена. За сваку од 3 трке се чува највреднији резултат у свакој години јуниорске категорије. Те 4 године дају 4 резултата који после служе за праћење прогреса током периода који се испитује. Поменуте 3 трке и по њихова 4 резултата се додају у почетни низи и то је крајњи резултат функције *scrape_best_times*.

```

def scrape_best_times(main_url, junior_cat, name, gender, group):
    all_string = [name, gender, group]
    fina_points_dict = {}
    all_fina_points_dict = {}

    for event in events:
        all_fina_points_dict[event] = [0, 0, 0, 0]

    for index, year in enumerate(junior_cat):
        url = f"{main_url}&pbest={year}"
        r = requests.get(url)
        soup = BeautifulSoup(r.content, 'html.parser')

        race_table = soup.find('table', class_='athleteBest')
        table_rows = race_table.find_all('tr')

        for event in events:
            fina_points_dict[event] = 0

        for row in table_rows:
            event_cell = row.find('td', class_='event')
            course_cell = row.find('td', class_='course')

```

```

fina_point_cell = row.find('td', class_='code')

if event_cell is None or course_cell is None or fina_point_cell is
    None:
        continue

event = event_cell.text.strip()
course = course_cell.text.strip()
fina_point = fina_point_cell.text.strip()

if event in events and course == '50m':
    fina_points_dict[event] = fina_point
    all_fina_points_dict[event][index] = int(fina_point)

top_3_races = get_top_3_races(all_fina_points_dict)
print(f"\nTop 3 {name}'s races:")
for race, total_points in top_3_races.items():
    print(f"{race}: {total_points}")

#Turning it into an array so I can make a CSV file
for race, points in top_3_races.items():
    all_string.append(f"{race}")
    for i in range(1, 5):
        all_string.append(f"{points[i-1]}")

return all_string

```

2.2.3 Функција *write_to_csv*

На крају се финални низ од 21 елемената складишти у CSV фајл уз помоћ функције *write_to_csv* и процес се понавља за следећег пливача. Након што `for` петља прође кроз све итерације добија се CSV фајл са подацима за све пливаче и пливачице.

```

def write_to_csv(filename, data):
    with open(filename, 'a', newline='') as csvfile:
        writer = csv.writer(csvfile)
        #writer.writerow(header) This line is needed for the first run
        writer.writerow(data)

```

2.2.4 Сређивање базе

Рачун у коду који служи за селектовање 3 трке одабир ради на основу највећег просека током јуниорске категорије. Тај рачун има једну оправдану ману која је ручно решена. Та мана је могућности резултата од 0 поена у некој од 4 године. То прави проблем, посебно ако се нађе у другој или трећој години јуниорске категорије, зато што при мерењу прогреса представља нагли пад или раст, док је у ствари нешто друго. Проблем постоји зато што на почетку рачуна све 4 године трке имају вредност од 0 FINA поена. Програм ради тако што пролази кроз све могуће поене у једној години и ако нађе веће постави их као највеће те године. Већина пливача у истраживању су чланови репрезентације и способни су да отпливају квалитетне резултате. То значи да пливач има 0 поена само ако ту трку није пливао те године. Ова мана је оправдана зато што се пливачима

деси да неке године не пливају одређену трку, посебно ако им она није најбоља него друга или трећа. Главни разлог за то је што пливачи у млађим годинама експериментирају са тркама да би одредили које им одговарају. Такође су јуниорске категорије великог број пливача биле захваћене периодом COVID-19 вируса који је онемогућио одржавање такмичења, а самим тим и испливање резултата. Чланови истраживања су претежно мушког пола, зато су из узорка, ради релевантности резултата, избачени сви који имају године у којима је вредност поена 0. Са друге стране девојака има мање и избацивање свих пливачица са 0 поена не би било вредно зато што би се изгубило $\sim 15\%$ женског дела узорка. Зато је за 3 девојке које имају по 1 трку са 0 поена измењена та вредност тако да прати прогрес осталих година. Девојке које су имале више од једне трке са 0 поена избачене су из узорка. Овако су првобитно ¹ изгледали резултати напоменуте 3 девојке:

Name	Gender	Group	Length3	Stroke3	I	II	III	IV
Дуња Остојић	2	2	200	Freestyle	558	0	609	557
Дуња Остојић	2	3	200	Freestyle	521	558	0	609
Сара Блитва	2	2	100	Freestyle	0	552	541	551
Сара Блитва	2	3	50	Freestyle	0	500	552	500
Јелена Јечански	2	3	50	Backstroke	0	646	669	701

Табела 1: Неизмењени редови пливачица које су имале 0 поена у једној години јуниорске категорије.

Следећи рачуни редом представљу на који начин су добијене вредности које ће заменити 0 у табели:

$$\frac{\frac{1}{2}(609 - 558) + (557 - 609)}{2} \approx -13, \text{ што значи да вредност } 558 - 13 = 545 \text{ мења вредност } 0.$$

$$\frac{(558 - 521) + \frac{1}{2}(609 - 558)}{2} \approx 31, \text{ што значи да вредност } 558 + 31 = 589 \text{ мења вредност } 0.$$

$$\frac{(541 - 552) + (551 - 541)}{2} \approx 0, \text{ што значи да вредност } 552 - 0 = 552 \text{ мења вредност } 0.$$

$$\frac{(552 - 500) + (500 - 552)}{2} = 0, \text{ што значи да вредност } 500 - 0 = 500 \text{ мења вредност } 0.$$

$$\frac{(669 - 646) + (701 - 669)}{2} \approx 28, \text{ што значи да вредност } 646 - 28 = 618 \text{ мења вредност } 0.$$

¹Ово није попуни, него модификовани приказ свих променљивих зато што приказује само колоне које се мењају.

Крајњи резултат је:

Name	Gender	Group	Length3	Stroke3	I	II	III	IV
Дуња Остојић	2	2	200	Freestyle	558	545	609	557
Дуња Остојић	2	3	200	Freestyle	521	558	589	609
Сара Блитва	2	2	100	Freestyle	552	552	541	551
Сара Блитва	2	3	50	Freestyle	500	500	552	500
Јелена Јечански	2	3	50	Backstroke	618	646	669	701

Табела 2: Измењени редови пливачица које су имале 0 поена у једној години јуниорске категорије.

3 Статистички тестови

3.1 Подела типова података

Пре него што се почне са статистичким тестовима потребно је прво проверити особине променљивих на којима ће се они радити. Прво се посматра класификација променљивих на основу њиховог типа података. Резултати ове поделе су корисни зато што одређени статистички тестови захтевају одређене типове података. Ова класификација дели променљиве на:

- **Категоријалне променљиве:** назив пливача (Name), пол пливача (Gender), група пливача (Group), дужина трке (Length), стил трке (Stroke).
- **Нумеричке променљиве:** FINA поени.

Напомена: Променљиве попут дужина трке, стил трке и FINA поени су напоменуте само једном у класификацији али постоји више њихових инстанци. Тако у бази постоје колоне Length1, Length2 и Length3, као и колоне Stroke1, Stroke2 и Stroke3. Постоје по 3 колоне за ове променљиве зато што се тестови раде на 3 трке. Свака од њих се посматра у период од 4 године па зато постоји 12 колона FINA поена који су представљени комбинацијом арапских и римских бројева. Бројеви од 1 до 3 представљају трке, док знакови од I, II, III и IV представљају годину јуниорске категорије током којих је испливана трка. За сваког пливача посматрано је по 3 трке. То је урађено да се минимизују ефекти случајности и специфичних услова појединачних трка, пружајући прецизнију процену напретка. Такође, коришћењем већег броја података за сваког пливача повећава се статистичка моћ анализе, што омогућава поузданије закључке о утицају промена у старосним категоријама.

3.2 Провера нормалне расподеле променљивих

Тестови се раде само на FINA поенима и из тога разлога се само код њих проверава нормална расподела података. За проверу нормалности података користи се Shapiro-Wilk тест на нивоу значајности од 0.05. Тестира се нулта хипотеза (null hypothesis) која тврди да подаци имају нормалну расподелу против алтернативне хипотезе (alternative hypothesis) која тврди супротно, тј. да подаци немају нормалну расподелу. Ако је p вредност мања од 0.05, нулта хипотеза се одбацује, а у супротном случају ако је p вредност већа од 0.05, нулта хипотеза се не одбацује, што значи да не постоје докази против претпоставке нормалности. Добијени су следећи резултати:

	1I	1II	1III	1IV	2I	2II	2III	2IV	3I	3II	3III	3IV
$p =$	0.83	0.89	0.60	0.22	0.82	0.14	0.69	0.59	0.58	0.18	0.37	0.23

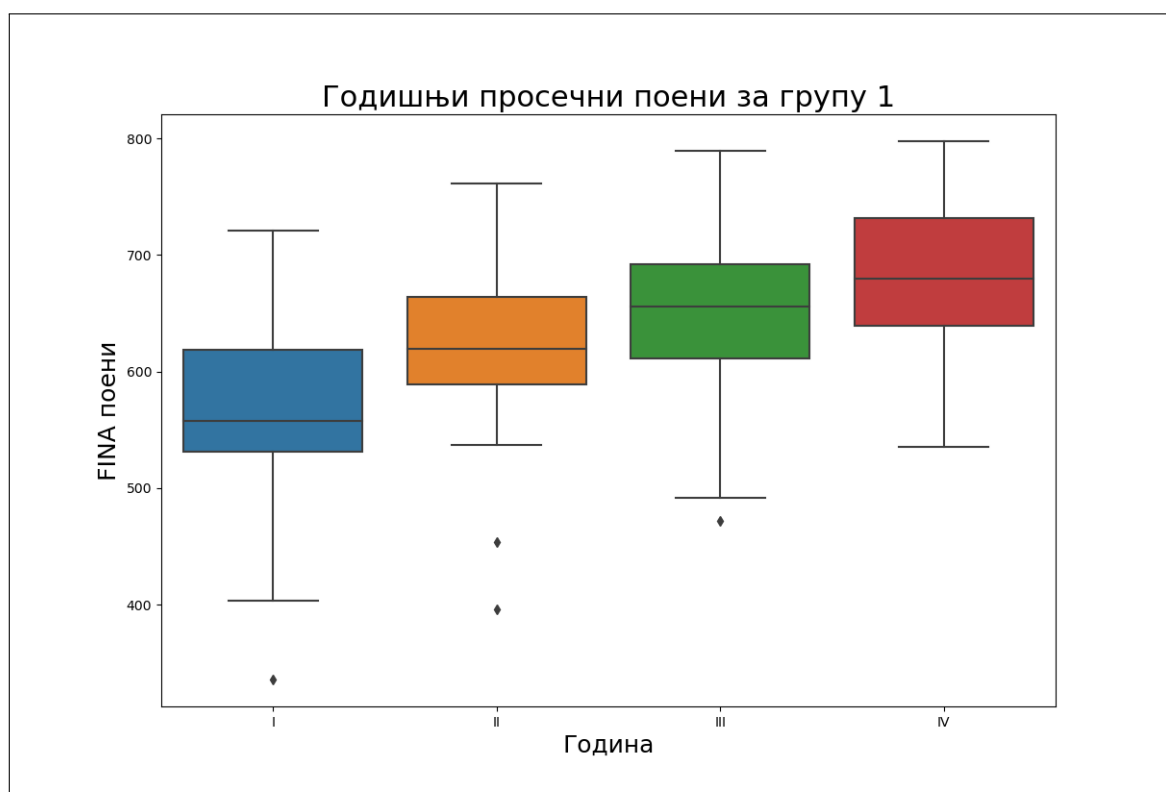
Табела 3: p вредности променљивих на којима се раде статистички тестови.

Из табеле се закључује да све колоне са FINA поенима имају p вредност већу од 0.05, што значи да се прихвата нулта хипотеза и да све променљиве на којима се раде статистички тестови имају нормалну расподелу.

3.3 Дескриптивна статистика

Дескриптивна статистика служи да прикаже основне карактеристике скупа података и на тај начин ствара почетно разумевање података. Ово истраживање пореди 3 скупа са дугачким називима и стога се за даљи текст уводи следећа терминологија: група 1, група 2 и група 3. Називи редом представљају мушку јуниорску категорију током које момци имају од 15 до 18 година, женску јуниорску категорију током које девојке имају од 15 до 18 година и женску јуниорску категорију током које девојке имају од 14 до 17 година.

3.3.1 Дескриптивна статистика групе 1



Слика 1: Годишњи просечни поени током јуниорске категорије за момке 15-18.

	Просек	Q1	Медијана	Q3	Range	IQR
I	562.9122	531.2500	557.5000	618.4166	385.0000	87.1666
II	622.0000	589.2500	619.6666	664.4166	365.3333	75.1666
III	653.5087	610.8333	655.8333	691.7500	317.6666	80.9166
IV	676.0526	639.2500	680.0000	731.3333	262.6666	92.0833

Табела 4: Дескриптивна статистика за момке 15-18.

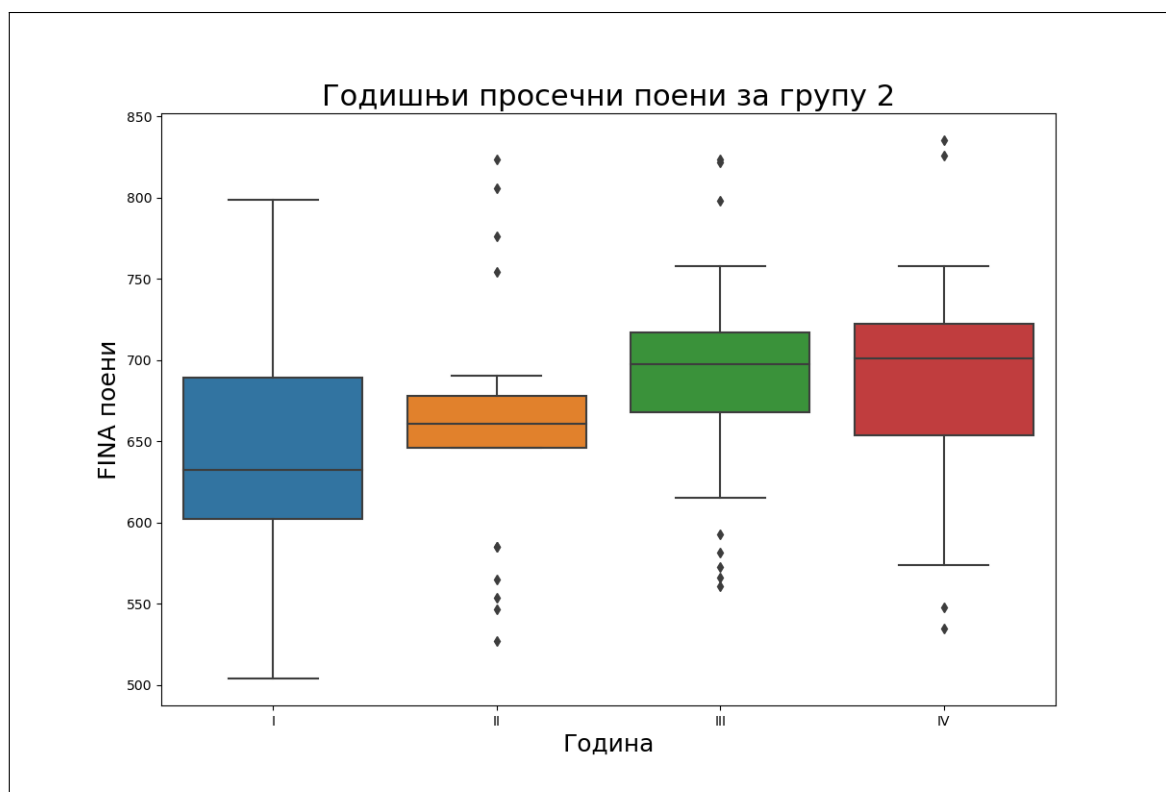
Са претходне слике и графикана могу се извести следећи закључци:

- Просечни поени константно расту од године I (562,91) до године IV (676,05). У свакој години, медијана је близу просека, што сугерише релативно симетричну

дистрибуцију са благим десним репом, јер су просеци генерално виши од медијана. Десни реп представља постајање већих вредности које померају просек на десно. Постојање акцелераната који су раније ушли у пубертет може објаснити ову појаву.

- Опсег се смањује током времена, што сугерише конзистентније перформансе међу учесницима како напредују. То може бити објашњено иницијалном разликом између ретарданата и акцелераната у почетним годинама јуниорске категорије. Временом, када оба та гранична случаја достигну сличан ниво пубертетског развоја, та иницијална разлика се изнивелише. IQR показује одређену варијабилност, али остаје релативно стабилан, што указује на конзистентан распон међу кватилима.

3.3.2 Дескриптивна статистика групе 2



Слика 2: Годишњи просечни поени током јуниорске категорије за девојке 15-18.

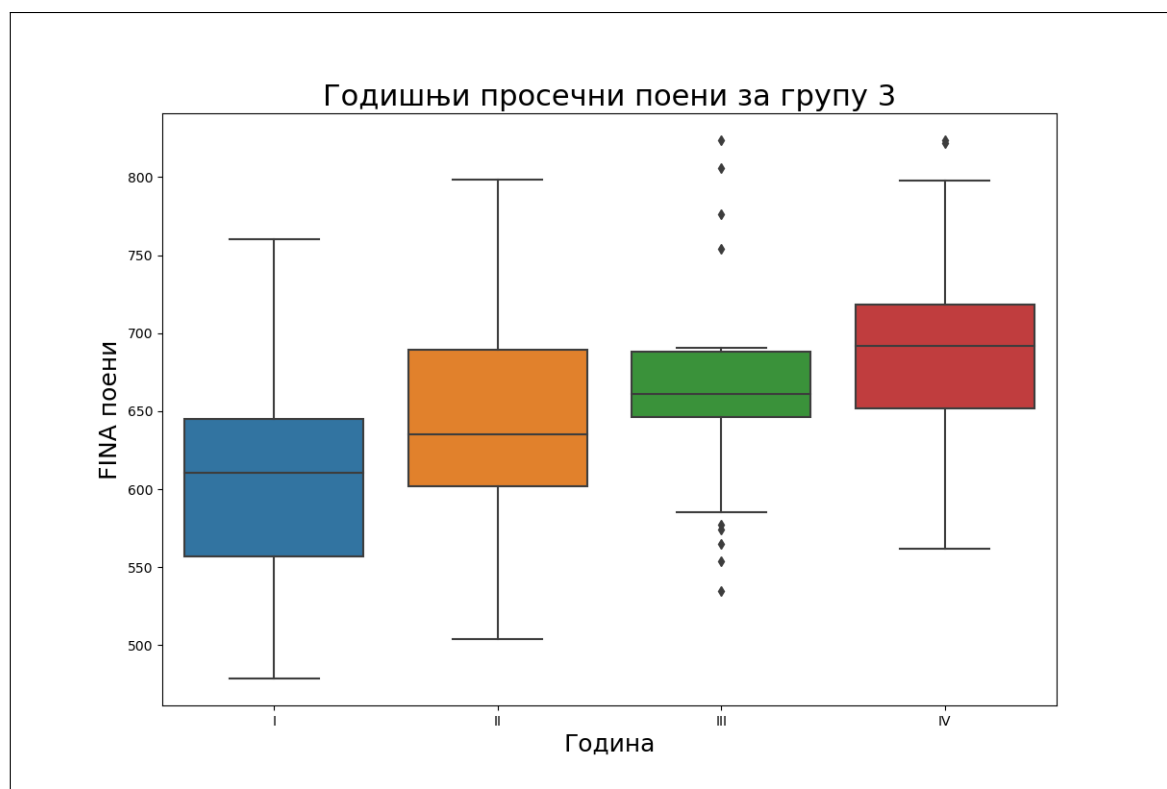
	Просек	Q1	Медијана	Q3	Range	IQR
I	644.0400	602.0000	632.6667	689.3333	294.6667	87.3333
II	660.2267	646.0000	661.0000	678.0000	296.3333	32.0000
III	688.8933	667.6667	697.6667	717.0000	262.6667	49.3333
IV	686.3067	653.6667	701.0000	722.3333	300.3333	68.6667

Табела 5: Дескриптивна статистика за девојке 15-18.

Са претходне слике и графикана могу се извести следећи закључци:

- Просечни поени такође расту од године I (644,04) до године IV (686,31). Тај раст је знатно мањи него код групе 1 и представља свега $\sim 37\%$ њега. Медијане су ближе просеку у већини година, што сугерише симетричнију дистрибуцију. То значи да не постоји пуно екстремних вредности који би утицале на просек тако да се он помери на једну или другу страну медијане.
- Опсег и IQR су мањи него у групи 1, осим за годину II где је IQR значајно мањи. Већа коцентрисаност резултата може се објаснити тиме што је код већине пливачица већ достигнут сличан ниво пубертетског развоја, што би значило да не постоји велик број граничних вредности.

3.3.3 Дескриптивна статистика групе 3



Слика 3: Годишњи просечни поени током јуниорске категорије за девојке 14-17.

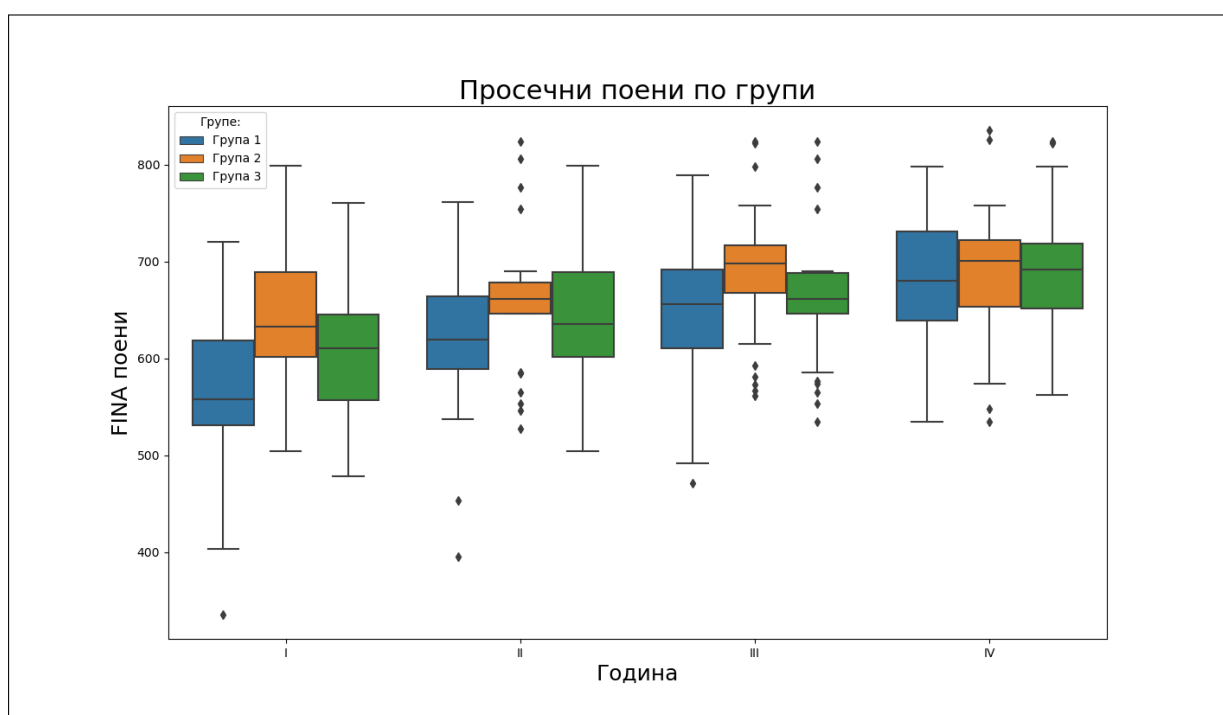
	Просек	Q1	Медијана	Q3	Range	IQR
I	606.6267	557.0000	610.3333	645.0000	281.3333	88.0000
II	646.4667	602.0000	635.3333	689.3333	294.6667	87.3333
III	661.3200	646.0000	661.0000	688.3333	289.0000	42.3333
IV	686.8933	651.6667	692.0000	718.3333	261.6667	66.6667

Табела 6: Дескриптивна статистика за девојке 14-17.

Са претходне слике и графикана могу се извести следећи закључци:

- Просечни поени такође расту од године I (606,63) до године IV (686,89). Тај раст је мањи него код групе 1 и представља $\sim 70\%$ њега. Иако се тај проценат на први поглед не чини довољно великим, он је душло већи него проценат разлике групе 2 у односу на групу 1. Ово је први указатељ на то да је прогрес између групе 1 сличнији прогресу групе 3 него прогресу групе 2.
- Опсег и IQR показују сличан образац као у групи 1, што указује на конзистентан распон података. Година III има најмањи IQR (42,33), што указује на мању варијабилност међу средњих 50 % учесника.

3.3.4 Дескриптивна статистика свих група заједно



Слика 4: Годишњи просечни поени током јуниорксе категорије за све групе.

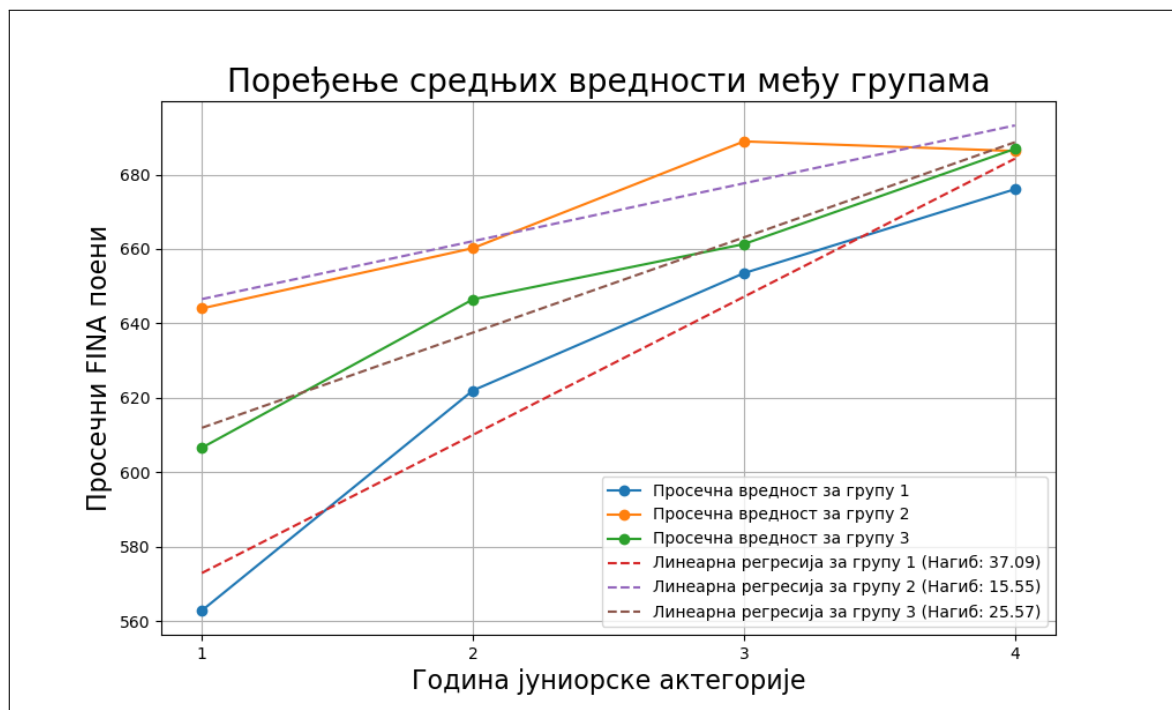
Са претходне слике изводи се следећи закључак:

- Визуелна репрезентација свих група заједно приказује како кроз период јуниорске категорије група 1 и група 3 теже ка групи 2. Старе јуниорске категорије су на почетку интервала коцентрисане око сличних вредности FINA поена, док је нова женска категорија фокусирана око знатно већих вредности FINA поена. Временом, како чланови групе 1 и 2 достигну финалне развоје пубертетског периода сустижу FINA поене групе 3.

3.4 Једнострука линеарна регресија

Први статистички метод који се користи је једнострука линеарна регресија која претпоставља линеарну везу између године јуниорске категорије и просечног

броја FINA поена у тој години [2]. Приоритет ове методе у овом истраживању није на проналажењу што боље линеарне везе, него на нагибу праве која је представља. Нагиб поменутих криве у односу на x осу координатног система се може посматрати као брзина прогреса током јуниорске категорије. Уколико се подаци поделе на 3 дела у зависности од вредности променљиве "група пливача" могу се извући прогреси за сваку од три групе јуниора, који се даље могу упоређивати. Након поделе свака група има иситх 12 колона које представљају FINA поене за 3 различите трке у периоду од 4 године. За поређење су узете просечне вредности поена 3 трке у истој години категорије, а затим је пронађен нагиб криве линеарне регресије преостале 4 тачке. То је одређено за сваку групу и добијени су следећи резултати:



Слика 5: Поређење прогреса у 3 различите групе јуниорских пливача.

Нагиби кривих линеарне регресије означавају брзину прогреса група: група 1 (плава линија) има нагиб 37.09° , група 2 (наранџаста линија) 15.55° , а група 3 (зелена линија) 25.57° . Посматрајући ову метрику, закључујемо да група 3 остварује приближно $\sim 69\%$ прогреса групе 1, што је знатно већи проценат у поређењу са $\sim 42\%$ прогреса групе 2 у односу на групу 1. То указује на већу сличност између група 1 и 3 него између група 1 и 2.

На слици 1 може се приметити да дужи које повезују године групе 3 ближе прате успоне и падове дужи групе 1, него што то успевају дужи групе 2. Дужи група 1 и 3 су распоређене тако да прва има највећи нагиб, а затим следеће две нешто мањи. Са друге стране, једина сличност између групе 1 и 2 је што период између претпоследње и последње године представља интервал са најмање напретка. Ово још једном показује да већу сличност имају прогреси група 1 и група 3 него група 1 и група 2.

3.5 T тест

T тест је врста параметарског теста који утврђује да ли постоји статистички значајна разлика између средњих вредности два скупа [2]. У овом истраживању ти скупови су FINA поени из исте године јуниорске категорије код различитих група. T тест са овим скуповима показује да ли су пливачи различитих група у истом тренутку јуниорске категорије на истом нивоу. За овај тест је опет потребна подела у три различите базе где свака представља једну од група. Иако је у делу 3.1 показано да у почетној бази све променљиве имају нормалну расподелу, то не значи да ће те исте променљиве подељене у три базе опет поседовати ту особину. Из тога разлога се понавља Shapiro-Wilk тест за сваку променљиву за све 3 базе. Добијене су следеће p вредности:

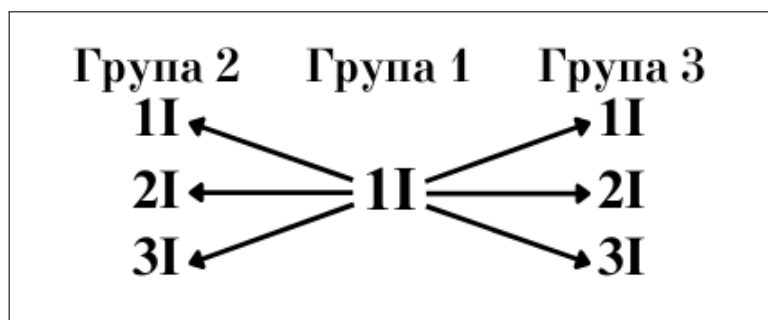
	Група 1	Група 2	Група 3
1I	0.4094	0.3227	0.9728
1II	0.3277	0.4512	0.6607
1III	0.9393	0.1841	0.2549
1IV	0.1935	0.2432	0.3301
2I	0.3847	0.7017	0.9600
2II	0.0160	0.2441	0.6834
2III	0.5426	0.3893	0.1839
2IV	0.6709	0.3207	0.8327
3I	0.5950	0.7772	0.7768
3II	0.3316	0.0452	0.9459
3III	0.3698	0.4482	0.0802
3IV	0.0687	0.5031	0.5721

Табела 7: p вредности променљивих свих група над којима се раде статистички тестови.

Све променљиве сем друге трке у другој години јуниорске категорије код групе 1 и треће трке друге године јуниорске категорије код групе 2 имају нормалну расподелу података. Зато се код тих променљивих уместо T теста ради непараметарски Mann-Whitney U тест [2]. Код T теста се тестира нулта хипотеза (null hypothesis) која тврди да нема разлике у средњим вредностима између две групе против алтернативне хипотезе (alternative hypothesis) која тврди да постоји разлика у средњим вредностима. Ако је p вредност мања од 0.05, нулта хипотеза се одбацује, што значи да постоји статистички значајна разлика између средњих вредности две групе. Ако је p вредност већа од 0.05, нулта хипотеза се не одбацује, што значи да нема статистички значајне разлике између средњих вредности две групе. Mann-Whitney U тест тестира исте хипотезе као T тест, али то ради на подацима који не прате нормалну расподелу.

Тестови се извршавају тако што се FINA поени једне трке у одређеној години групе 1 пореде са FINA поенима све 3 трке у истој тој години групе 2 и 3. За сваку од 12 променљивих из групе 1 се ради по том принципу, што на крају даје 72 теста.

Визуелни приказ система за тестирање:



Слика 6: Појашњење комбинација параметара на којима се ради Т и Mann-Whitney U тестови.

За Т и Mann-Whitney U тестове између група 1 и 2 добијене су следеће p вредности:

	1I	1II	1III	1IV	2I	2II	2III	2IV	3I	3II	3III	3IV
1I	0.0005				0.0094				0.0197			
1II		0.1407				0.5290				0.8246		
1III			0.0882				0.6349				0.6781	
1IV				0.7713				0.2236				0.0524
2I	0.0000				0.0006				0.0013			
2II		0.0321				0.2640				0.7096		
2III			0.0019				0.0635				0.3216	
2IV				0.1459				0.6069				0.8655
3I	0.0000				0.0000				0.0000			
3II		0.0002				0.0039				0.0267		
3III			0.0001				0.0059				0.0522	
3IV				0.0038				0.0477				0.1756

Табела 8: p вредности Т и Mann-Whitney U тестова између група 1 и 2.

За Т и Mann-Whitney U тестове између група 1 и 3 добијене су следеће p вредности:

	1I	1II	1III	1IV	2I	2II	2III	2IV	3I	3II	3III	3IV
1I	0.0590				0.3595				0.4622			
1II		0.4100				0.9339				0.7149		
1III			0.9097				0.3924				0.0883	
1IV				0.9225				0.2168				0.0381
2I	0.0051				0.0674				0.0993			
2II		0.0079				0.0508				0.2242		
2III			0.1327				0.6095				0.7342	
2IV				0.1057				0.5729				0.7691
3I	0.0002				0.0046				0.0078			
3II		0.0018				0.0316				0.1347		
3III			0.0155				0.1465				0.5218	
3IV				0.0025				0.0375				0.2136

Табела 9: p вредности Т и Mann-Whitney U тестова између група 1 и 3.

Може се приметити да између група 1 и 2 постоји 18 статистички значајних разлика док између група 1 и 3 постоји 10 статистички значајних разлика. То још

једном показује да прогрес групе 1 има више сличности са прогресом групе 3, него се прогресом групе 2.

3.6 ANOVA

ANOVA, или анализа варијансе, је статистичка метода која се користи за упоређивање средњих вредности између три или више група како би се утврдило да ли постоје статистички значајне разлике међу њима [2]. У овом истраживању те групе су јуниорске категорије а тестови ће бити одређени на FINA поенима. Предуслови ANOVA су нормална расподела међу променљивим као и њихова хомоскедастичност. Нормална расподела је проверена у области 3.5 и на основу тих резултата се не раде ANOVA тестови на променљивим 2II и 3II. За остале променљиве се хомоскедастичност проверава уз помоћ Levene теста. Levene тест се користи за тестирање хомогености варијанси између две или више група података. Тестира се нулта хипотеза (null hypothesis) која тврди да су варијансе свих група једнаке против алтернативне хипотезе (alternative hypothesis) која тврди да барем једна група има варијансу која се разликује од осталих. Ако је p вредност мања од 0.05, нулта хипотеза се одбацује, што значи да постоји статистички значајна разлика између варијанси група. Ако је p вредност већа од 0.05, нулта хипотеза се не одбацује, што значи да нема статистички значајне разлике између варијанси група. Levene тест је дао следеће p вредности:

	1I	1II	1III	1IV	2I	2II	2IV	3I	3II	3IV
$p =$	0.7710	0.8251	0.8904	0.9232	0.9328	0.9323	0.9765	0.9536	0.8902	0.9417

Табела 10: p вредности Levene теста.

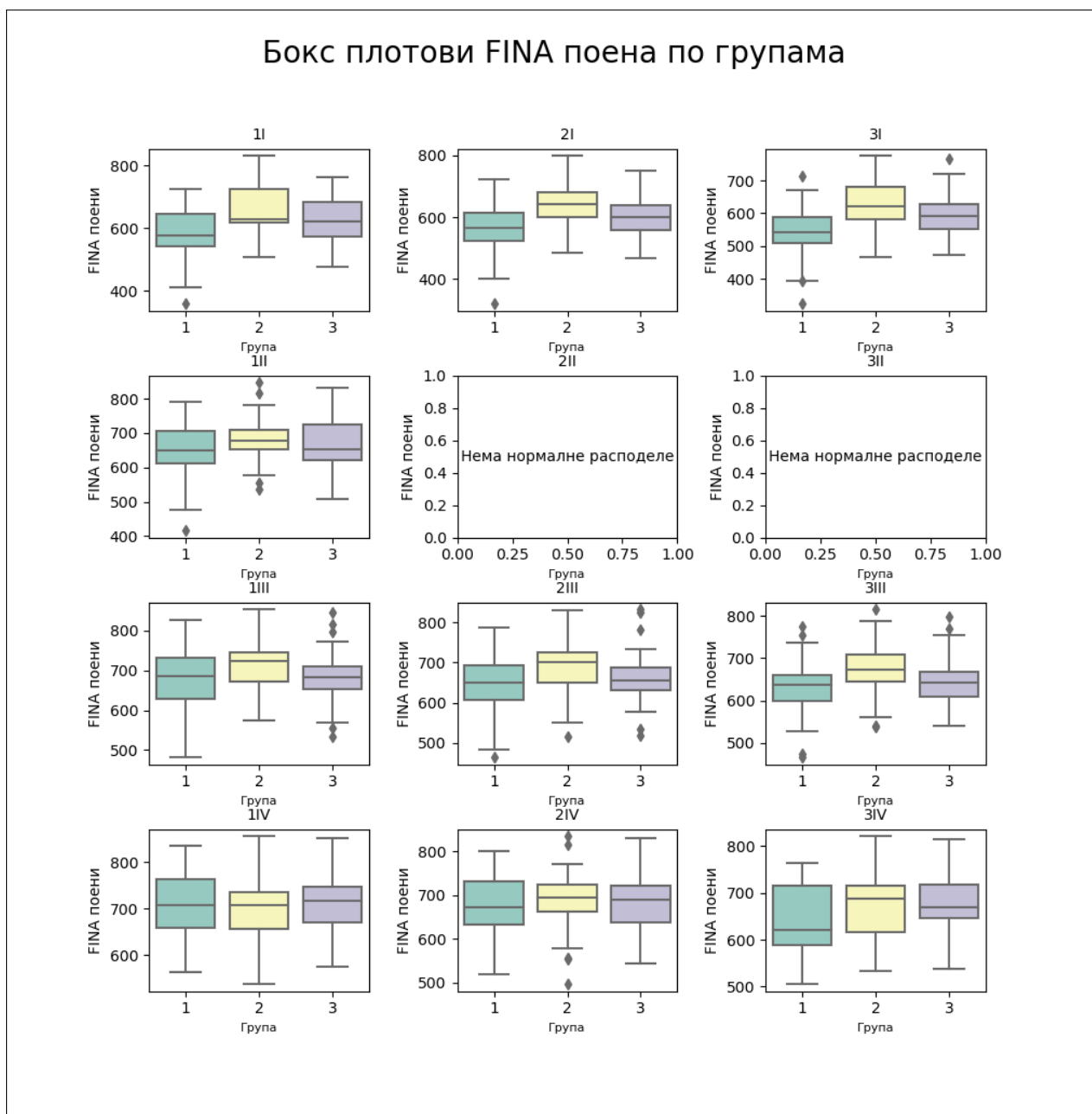
Levene тест је за све променљиве добио p вредности већу од 0.05, што значи да су све дисперзије једнаке и може се наставити са тестом за ANOVA. Код ANOVA теста нулта хипотеза (null hypothesis) тврди да су све групе из узорка међусобно једнаке, тј. да не постоји статистички значајна разлика између њихових средњих вредности. Алтернативна хипотеза (alternative hypothesis) тврди да постоји барем једна група која се разликује од осталих. Ако је p вредност мања од 0.05, нулта хипотеза се одбацује, што значи да постоји статистички значајна разлика између средњих вредности најмање две групе. У супротном случају, ако је p вредност већа од 0.05, нулта хипотеза се не одбацује, што значи да нема довољно доказа да се утврди да постоје разлике између група. ANOVA тест је дао следеће p вредности:

	1I	1II	1III	1IV	2I	2II	2IV	3I	3II	3IV
$p =$	0.0010	0.3234	0.1959	0.9601	0.0016	0.1745	0.8141	0.0001	0.1422	0.2976

Табела 11: p вредности резултата ANOVA теста.

ANOVA тест је за 3 променљиве добио p вредности мању од 0.05, што значи да међу њима постоје статистички значајне разлике у средњим вредностима међу неким групама. Да би се тачно утврдило које су то групе користи се Tukey HSD (Honestly Significant Difference) тест [2]. Нулта хипотеза (null hypothesis) тврди да не постоји статистички значајна разлика између парова група. Алтернативна

хипотеза (alternative hypothesis) тврди да постоји статистички значајна разлика између парова група. Ако је p вредност мања од 0.05, нулта хипотеза се одбацује за тај пар група, што значи да постоји статистички значајна разлика између средњих вредности тих двеју група. Ако је p вредност већа од 0.05, нулта хипотеза се не одбацује, што значи да не постоји статистички значајна разлика између средњих вредности тих двеју група. Пре него што се уради Tukey HSD тест приказаће се скуп графикона који визуелно откривају где се налазе значајне разлике између парова група:



Слика 7: Појашњење комбинација тестова који се раде.

Најзначајне разлике међу групама се могу наћи код променљивих које представљају 1. годину у јуниорској категорији. То статистички потврђују резултати Tukey HSD теста који даје следеће p вредности:

Пар група	1I	2I	3I
1 - 2	0.0007	0.0905	0.1207
1 - 3	0.1447	0.3552	0.8038
2 - 3	0.1827	0.7766	0.4341

Табела 12: p вредности резултата Tukey HSD теста.

Tukey HSD тест је за променљиву 1I група 1 и 2 добио p вредност 0.0007, што значи да постоји статистички значајна разлика међу њиховим средњим вредностима. Такође се може приметити да је p вредност за све променљиве увек најмања међу групама 1 и 2, што опет указује на већу сличност у прогресима међу групама 1 и 3.

4 Закључак

Истраживање је показало да промене у старосним категоријама за јуниорске пливаче имају значајан утицај на напредак спортиста. Дескриптивна статистика указује да просечни FINA поени расту од прве до четврте године, при чему је раст групе 3 (жене 14-17) око 70° раста групе 1 (мушкарци 15-18). Иако се овај проценат чини малим, он је двоструко већи у поређењу са групом 2 (жене 15-18). Ово сугерише да је напредак групе 3 сличнији групи 1 него групи 2.

Визуелна репрезентација приказује како се кроз јуниорски период групе 1 и 3 приближавају групи 2. Старе јуниорске категорије почињу са сличним FINA поенима, док нова женска категорија стартује са знатно већим вредностима. Временом, како чланови група 1 и 3 достижу финалне фазе пубертета, њихови FINA поени се приближавају онима из групе 2.

Линијска регресија показује нагибе од 37.09° , 15.55° , и 25.57° за групе 1, 2 и 3 респективно. Ово указује да група 3 остварује око 69° прогреса групе 1, што је знатно већи проценат од 42° колико износи прогрес групе 2 у односу на групу 1. На слици 1, дужи групе 3 ближе прате успоне и падове дужи групе 1 него групе 2, потврђујући већу сличност између група 1 и 3.

Статистички тестови, Т тест и Mann-Whitney U тест, показују 18 значајних разлика између група 1 и 2, те 10 значајних разлика између група 1 и 3. Ово додатно потврђује да је прогрес групе 1 сличнији прогресу групе 3 него групи 2.

ANOVA тест је показао статистички значајне разлике међу групама, посебно у првој години јуниорски категорија, што је потврдио Tukey HSD тест. Најзначајније разлике су између група 1 и 2, што још једном указује на већу сличност између група 1 и 3.

Ови резултати сугеришу да је стара женска јуниорска категорија (14-17 година) ближа мушкој јуниорској категорији (15-18 година) у смислу напретка него што је то нова категорија (15-18 година). Ова промена може имати импликације на тренинг и такмичарске стратегије за младе спортисте.

Литература

- [1] Paulo Francisco de Almeida-Neto, Dihogo Gama de Matos, Vanessa Carla Monteiro Pinto, Paulo Moreira Silva Dantas, Tatianny de Macêdo Cesário, Luíz Felipe da Silva, Alexandre Bulhões-Correia, Felipe José Aidar, and Breno Guilherme de Araújo Tinôco Cabral. "Can the Neuromuscular Performance of Young Athletes Be Influenced by Hormone Levels and Different Stages of Puberty?" *International Journal of Environmental Research and Public Health*, vol. 17, no. 16, pp. 5637, 2020. <https://doi.org/10.3390/ijerph17165637>
- [2] Загорка Лозанов-Црвенковић. *Сџаџисџика*. Универзитет у Новом Саду, Природно Математички Факултет, Нови Сад, 2012.

Кратка биографија

Никола Попадић је рођен 6. јуна 2003. године у Новом Саду. Основну школу „Жарко Зрењанин“ завршио је са Вуковом дипломом. Током првих шест разреда основне школе, додатно је похађао школу за математику „Архимедес“. По завршетку основне школе, уписао се у Гимназију „Јован Јовановић Змај“ на смер „Ученици са посебним способностима за рачунарство и информатику“, коју је такође завршио са Вуковом дипломом. Након тога, уписао је основне студије на Природно-математичком факултету у Новом Саду, смер примењена математика, модул аналитика података и статистика, где је све испите положио са просечном оценом 9,60.

**УНИВЕРЗИТЕТ У НОВОМ САДУ
ПРИРОДНО-МАТЕМАТИЧКИ ФАКУЛТЕТ
КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА**

Редни број:

РБР

Идентификациони број:

ИБР

Тип документације: Монографска документација

ТД

Тип записа: Текстуални штампани материјал

ТЗ

Врста рада: Дипломски рад

ВР

Аутор: Никола Попадић

АУ

Ментор: Др Загорка Лозанов-Црвенковић

МН

Наслов рада: Процена развоја јуниорских пливача у новим и старим старосним категоријама: Статистичка анализа

НР

Језик публикације: српски

ЈП

Језик извода: српски

ЈИ

Земља публикавања: Република Србија

ЗП

Уже географско подручје: Војводина

УГП

Година: 2024

ГО

Издавач: Ауторски репринт

ИЗ

Место и адреса: Нови Сад, Трг Доситеја Обрадовића 4

МА

Физички опис рада: 3/23/0/12/7/0/0 (број поглавља/страна/лит. цитата/табела/слика/графика/прилога)

ФО

Научна област: Математика

НО

Научна дисциплина: Статистика

НД

Предметна одредница/Кључне речи: јуниорска категорија, пливање, статистика

ПО

УДК:

Чува се: Библиотека Департмана за математику и информатику, Нови Сад

ЧУ

Важна напомена:

ВН

Извод:

ИЗ

Датум прихватања теме од стране НН Већа:

ДП

Датум Одбране:

ДО

Чланови комисије:

Председник: проф Љиљана Цветковић, редовни професор, Природно-математички факултет, Универзитет у Новом Саду

Члан: проф. Мирјана Штрбоја, редовни професор, Природно-математички факултет, Универзитет у Новом Саду

Ментор: проф. Загорка Лозанов Црвенковић, редовни професор, Природно-математички факултет, Универзитет у Новом Саду

КО

**UNIVERSITY OF NOVI SAD
FACULTY OF SCIENCES
KEY DOCUMENTARY INFORMATION**

Accession number:

ANO

Identification number:

INO

Document type: Monograph type

DT

Type of record: Printed text

TR

Content code: Thesis

CC

Author: Nikola Popadić

AU

Mentor: Dr. Zagorka Lozanov-Crvenković

MN

Title: Assessment of the Development of Junior Swimmers in New and Old Age Categories: A Statistical Analysis

TI

Language of the text: Serbian

LT

Language of the abstract: Serbian

LA

Country of publication: Republic of Serbia

CP

Locality of publication: Vojvodina

LP

Publication year: 2024

PY

Publisher: Author's reprint

PU

Publication place: Novi Sad, Trg Dositeja Obradovića 4

PP

Physical description: 3/23/0/12/7/0/0 (number of chapters/pages/bibliographic citations/tables/figures/graphics/appendices)

PD

Scientific field: Mathematics

SF

Scientific discipline: Statistics

SD

Subject / Keywords: junior category, swimming, statistics

SKW

UC:

Holding data: Library of the Department of Mathematics and Informatics, Novi Sad

HD

Note:

N

Abstract:

AB

Accepted by Scientific Board on:

DP

Defended:

DE

Thesis defend board:

President: Prof. Ljiljana Cvetković, Full Professor, Faculty of Sciences, University of Novi Sad

Member: Prof. Mirjana Štrboja, Full Professor, Faculty of Sciences, University of Novi Sad

Mentor: Prof. Zagorka Lozanov Crvenković, Full Professor, Faculty of Sciences, University of Novi Sad

DB