



VILNIAUS UNIVERSITETAS

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Tiesiniai modeliai

Laboratorinis darbas

Atliko: 3 kurso 2 grupės studentai:

Matas Amšiejus

Salvija Račkauskaitė

Sandra Macijauskaitė

Darbo vadovė: doc. dr. Rūta Levulienė

Vilnius, 2021

TURINYS

ĮVADAS.....	4
1. TIRIAMASIS STRAIPSNIS	5
1.1.Tema.....	5
1.2.Duomenys	5
1.3.Atlikta analizė	6
2. ATLIKTA ANALIZĖ REMIANTIS STRAIPSNIU.....	6
2.1.Bendra analizės eiga.....	6
2.2.Ūgiai	6
2.3.Masės	10
2.4.Sprintas.....	12
2.5.Vikrumas.....	13
2.6.Driblingas.....	16
2.7.Kamuolio kontrolė.....	17
2.8.Kamuolio smūgiavimas.....	19
2.9.Python programavimo kalba	20
IŠVADOS	21
ŠALTINIAI	22

IVADAS

Šiame laboratoriniame darbe siekiame ištirti jaunų futbolininkų sugebėjimų įtaką jų vėlesniai profesionaliai karjerai. Sieksime pritaikyti vienfaktorę dispersinę analizę. Laboratorinio darbo uždavinį įgyvendinti pasitelksime R, SAS ir Python programavimo kalbas. Tyrimo tikslas – išanalizuoti ir pateikti pasirinkto straipsnio spredimus.

1. TIRIAMASIS STRAIPSNIS

Straipsnis „The influence of speed abilities and technical skills in early adolescence on adult success in soccer: A long-term prospective analysis using ANOVA and SEM approaches“ pasirinktas iš „Official Journal of the American College of Sports Medicine“ žurnalo, 5 tomo. Straipsnio autoriai: Oliver Höner, Daniel Leyhr, Augustin Kelava. Išleistas 2017 metais rugpjūčio 14 dieną.

1.1.Tema

Straipsnyje analizuojama U12 futbolininkų sugebėjimai ir jų įtaka tolimesnei karjerai. Šiame tyrime buvo įvertinta ilgalaikė prognozuojama motorinių testų, žaidėjų greičio ir techninių įgūdžių reikšmė ankstyvoje paauglystėje.

1.2.Duomenys

Tyrimo imtį sudarė 14178 U12 žaidėjai iš Vokietijos talentų ugdymo programos. Buvo atlikti penki testai (sprinto, vikrumo, driblingo, kamuolio kontrolė, kamuolio smūgiavimas), o žaidėjų ūgis, masė ir santykinis amžius buvo įvertinti visoje šalyje atliekant diagnostiką 2004–2006 m. 2014–15 m. sezono metu žaidėjai buvo suskirstyti į profesionalus (n = 89), pusiau profesionalus (n = 913) arba neprofesionalus (n = 13176). Į pirmas tris geriausias lygas patekę futbolininkai buvo priskirti į APL 1 (suaugusiųjų futbolo lygis) grupę, į ketvirtą – penktą į APL 2, o visi kiti į APL 3. Motorinių testų bei ūgio ir svorio prognostinė svarba buvo nustatyta naudojant dispersinę analizę. Duomenys pateikti pagal 9 kriterijus:

1. Age – amžius, metai;
2. Spr – sprintas, sekundės;
3. Agi – vikrumas, sekundės;
4. Dri – driblingas, sekundės;
5. Bc – kamuolio valdymas, sekundės;
6. Sho – kamuolio smūgiavimas, taškai;
7. Hei – ūgis, centimetrai;
8. Wei – masė, kilogramai;
9. Apl – suaugusiojo veiklos lygis, 1-3 lygiai.

Prieš atliekant motorinius testus, buvo užregistruotas kiekvieno žaidėjo ūgis, svoris ir santykinis amžius (matuojamas pagal gimimo dieną per kalendorinius metus). Motorinių gebėjimų įvertinimus sudarė 5 testai, mažesnės vertės rezultatai rodė geresnius pasiekimus. Žaidėjai buvo išbandyti sprinto (20

m bėgimas), vikrumo ir driblingo (laikas slalomo trasoje atitinkamai be ir su kamuoliu), kamuolio valdymo (šešių perdavimų laikas prieš dvi priešingas sienas) ir kamuolio smūgiavimas (8 spyriai į tris skirtingus taikinius, kuriuos treneris įvertino pagal tikslumą ir greitį).

1.3. Atlikta analizė

Tyrime buvo atlikta vienfaktorė dispersinė analizė, kurią pakartojome ir gavome tuos pačius rezultatus, kuriuos aptarsime 2 skyriuje.

2. ATLIKTA ANALIZĖ REMIANTIS STRAIPSNIU

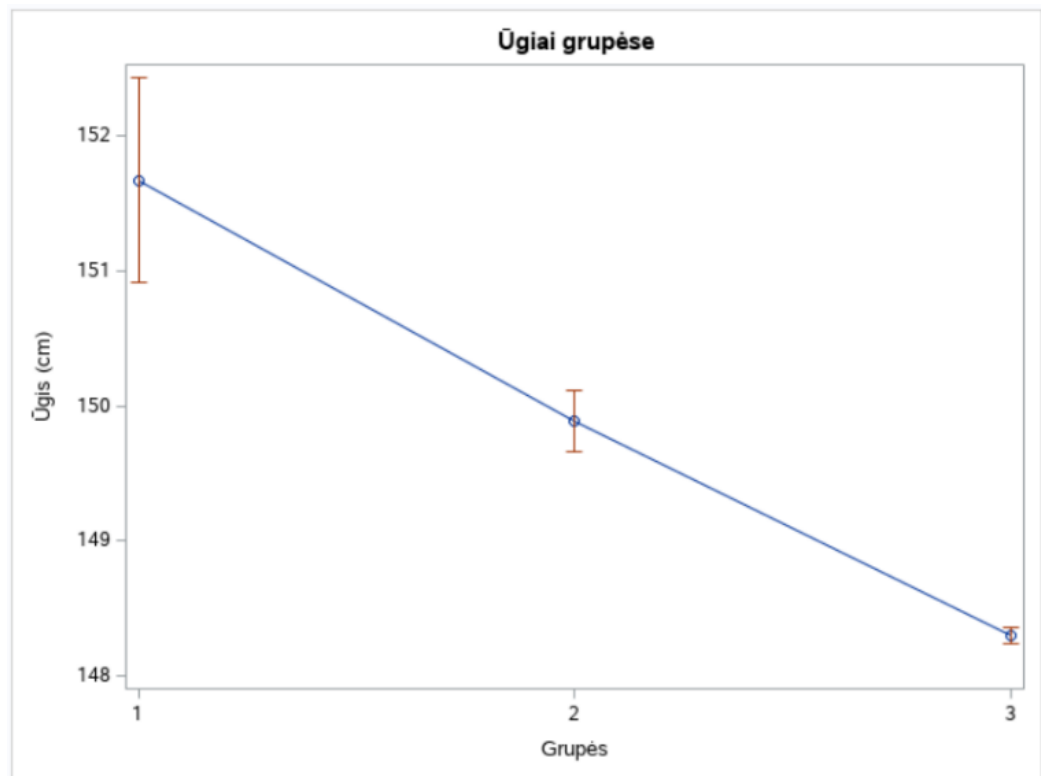
Atlikome vienfaktorę dispersinę analizę pagal straipsnį su SAS, R ir Python programavimo kalbomis.

2.1. Bendra analizės eiga

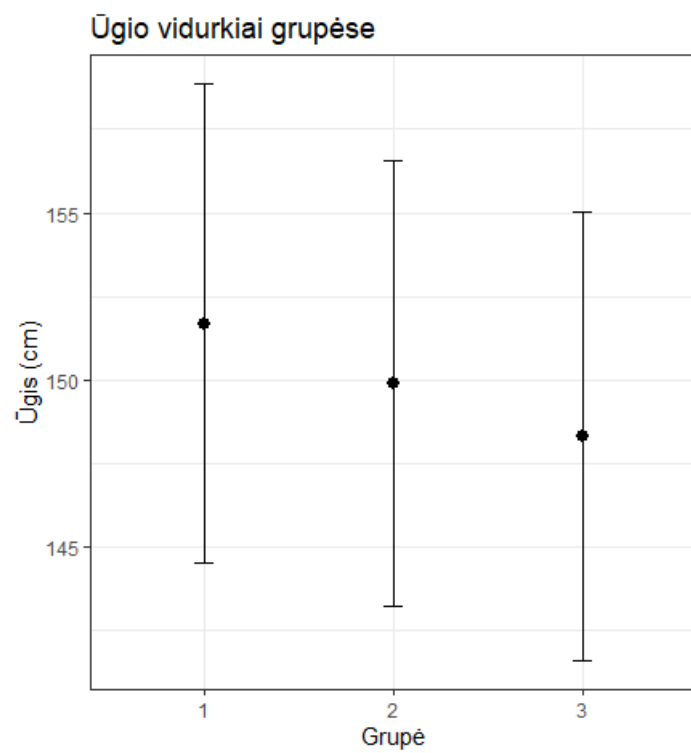
Pirmiausia nuskaitome duomenis iš S1File.txt. Tam, kad mūsų atliekamo tyrimo rezultatai sutaptų su straipsnio, mes perrenkame duomenis kaskart prieš tiriant naują kintamąjį. Atmetę nekorektiškas reikšmes, brėžiame vidurkių ir standartinių paklaidų grafikus (SAS) bei vidurkių ir standartinių nuokrypių grafikus (R). Tada atliekame dispersinę analizę (tolimesniame tyrime visur bus naudojamas reikšmingumo lygmuo $\alpha = 0,05$). Pirma tikriname nulinę hipotezę, kad dispersijos visose grupėse yra lygios, naudodami Levene testą. Jei nulinės hipotezės neatmetame (dispersijos lygios), tęsiame tyrimą. Jei ne, darbą tesiame naudodami Welch ANOVA. Tikriname nulinę hipotezę, kad vidurkiai visose grupėse yra lygūs. Jei taip, analizę baigiame. Jei ne, tikriname, kurie vidurkiai statistiškai reikšmingai skiriasi naudodami Tjūkio testą (SAS) bei Bonferoni (R). Patikriname, kokią dalį skirtumų lėmė lyga, o kokią – individualūs žaidėjų skirtumai (R-square).

2.2. Ūgiai

Pirma nusibraižome vidurkių ir standartinių paklaidų grafikus (SAS) bei vidurkių ir standartinių nuokrypių grafikus (R).



1 pav. Ūgio vidurkių ir standartinių paklaidų grafikas (SAS)



2 pav. Ūgio vidurkių ir standartinių nuokrypių grafikas (R)

Matome, kad ūgių vidurkiai tarp grupių mažėja. Taip pat mažėja ir standartinė paklaida, nes jos dydis smarkiai priklauso nuo imties dydžio (st. paklaida = st. nuokrypis / sqrt(imties dydis)). Tikriname nulinę hipotezę, kad dispersijos yra lygios.

The GLM Procedure					
Levene's Test for Homogeneity of Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
apl	2	3597.6	1798.8	0.39	0.6778
Error	13916	64369644	4625.6		

3 pav. Dispersijų lyginimo lentelė (SAS)

```

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group   2   0.1554 0.8561
      13916

```

4 pav. Dispersijų lyginimo lentelė (R)

Matome, kad p reikšmė = 0,6778 didesnė už reikšmingumo lygmenį, todėl nulinės hipotezės atmesti negalime (SAS). Su R gauname kitokią p reikšmę, tačiau išvada padarome tą pačią. Toliau tikriname hipotezę apie vidurkių lygybę.

Dependent Variable: hei					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3058.0106	1529.0053	33.89	<.0001
Error	13916	627853.1001	45.1174		
Corrected Total	13918	630911.1107			

5 pav. Vidurkių lyginimo lentelė (SAS)

```

      Df Sum Sq Mean Sq F value    Pr(>F)
apl      2    3058    1529.0    33.89 2.08e-15
Residuals 13916 627853     45.1

```

6 pav. Vidurkių lyginimo lentelė (R)

Gauname, kad p reikšmė yra mažiau už reikšmingumo lygmenį, todėl nulinę hipotezę atmetame, vidurkiai statistiškai reikšmingai skiriasi. Vadinasi ūgiai tarp grupių nėra vienodi. Tikriname, kurios grupės skiriasi.

Comparisons significant at the 0.05 level are indicated by ***.				
apl Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
1 - 2	1.78502	0.03495	3.53509	***
1 - 3	3.37478	1.70017	5.04940	***
2 - 1	-1.78502	-3.53509	-0.03495	***
2 - 3	1.58976	1.04502	2.13450	***
3 - 1	-3.37478	-5.04940	-1.70017	***
3 - 2	-1.58976	-2.13450	-1.04502	***

7 pav. Grupių tikrinimo lentelė (SAS)

1	2
2 0.05	-
3 7.0e-06	2.5e-11

8 pav. Grupių tikrinimo lentelė (R)

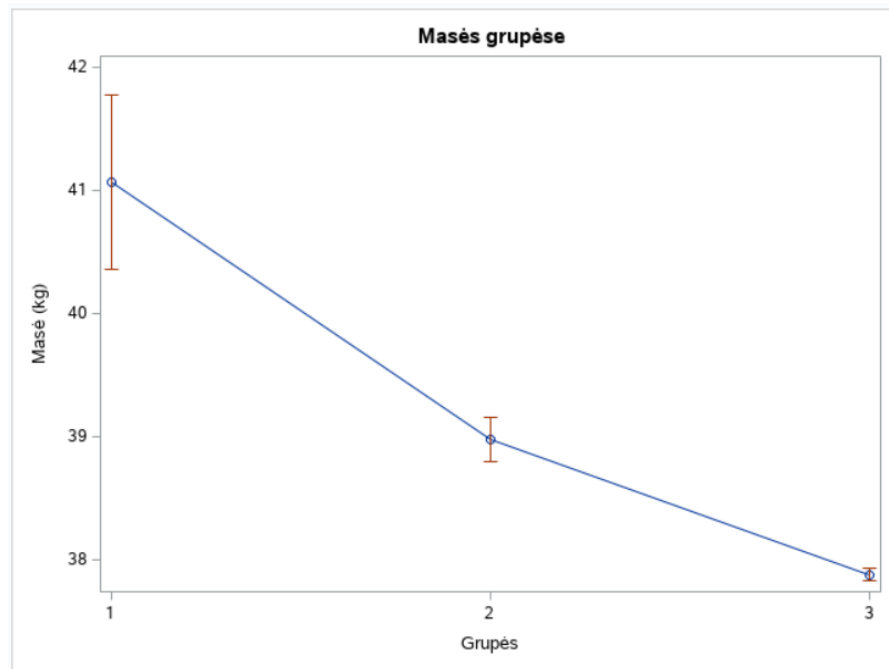
Pagal SAS lentelę (Tjūkio testą) matome, kad ūgiai statistiškai reikšmingai skiriasi tarp visų grupių (visur yra pažymėta ***), o pagal R (Bonferoni testą) pirmos ir antros grupės vidurkiai statistiškai reikšmingai nesiskiria su reikšmingumo lygmeniu 0,05. Patikrinkime R kvadrato reikšmes.

R-Square	Coeff Var	Root MSE	hei Mean
0.004847	4.525544	6.716945	148.4229

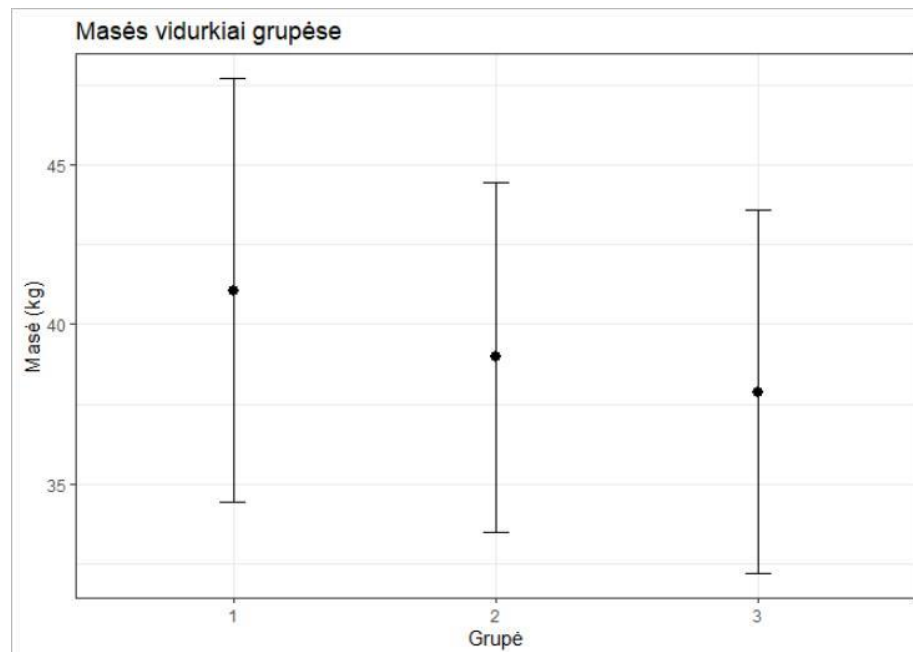
9 pav. R kvadrato lentelė

Gauname, kad tik < 0,01 % ūgių lėmė skirtingi faktoriai (futbolo žaidimo lygiai), o > 99 % – individualūs žaidėjų skirtumai.

2.3.Masės



10 pav. Masės vidurkių ir standartinių paklaidų grafikas (SAS)



11 pav. Masės vidurkių ir standartinių nuokrypių grafikas (R)

Masių vidurkiai vėl mažėja. Tikriname nulinę hipotezę, kad dispersijos yra lygios.

Levene's Test for Homogeneity of Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
apl	2	15992.3	7996.2	1.87	0.1540
Error	13908	59432515	4273.3		

12 pav. Dispersijų lyginimo lentelė (SAS)

```

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group   2   1.9222 0.1463
      13908

```

13 pav. Dispersijų lyginimo lentelė (R)

P reikšmė didesnė už reikšmingumo lygmenį, todėl nulinės hipotezės atmesti negalime. Tikriname hipotezę apie vidurkių lygybę.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1858.6766	929.3383	28.71	<.0001
Error	13908	450242.3471	32.3729		
Corrected Total	13910	452101.0237			

14 pav. Vidurkių lyginimo lentelė (SAS)

```

      Df Sum Sq Mean Sq F value    Pr(>F)
apl      2   1859    929.3    28.71 3.62e-13
Residuals 13908 450242     32.4

```

15 pav. Vidurkių lyginimo lentelė (R)

P reikšmė yra mažiau už reikšmingumo lygmenį, todėl nulinę hipotezę atmetame, vidurkiai statistiškai reikšmingai skiriasi. Vadinasi masės tarp grupių nėra vienodos. Tikriname, kurios grupės skiriasi.

Comparisons significant at the 0.05 level are indicated by ***.					
apl Comparison	Difference Between Means	Simultaneous 95% Confidence Limits			
1 - 2	2.08594	0.60350	3.56837	***	
1 - 3	3.18166	1.76315	4.60018	***	
2 - 1	-2.08594	-3.56837	-0.60350	***	
2 - 3	1.09573	0.63429	1.55717	***	
3 - 1	-3.18166	-4.60018	-1.76315	***	
3 - 2	-1.09573	-1.55717	-0.63429	***	

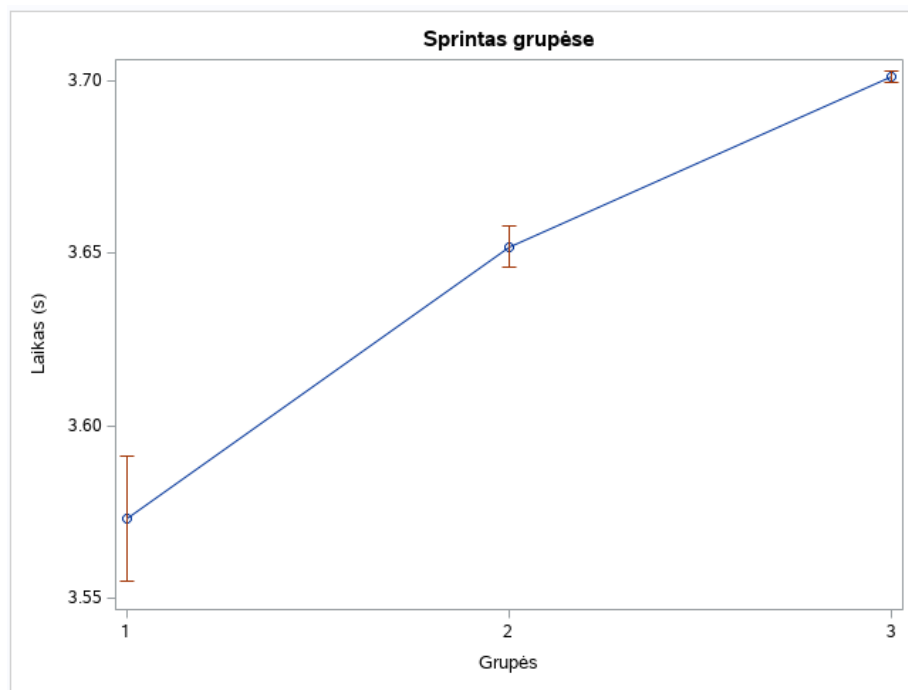
16 pav. Grupių lyginimo lentelė (SAS)

	1	2
2	0.0029	-
3	4.4e-07	8.0e-08

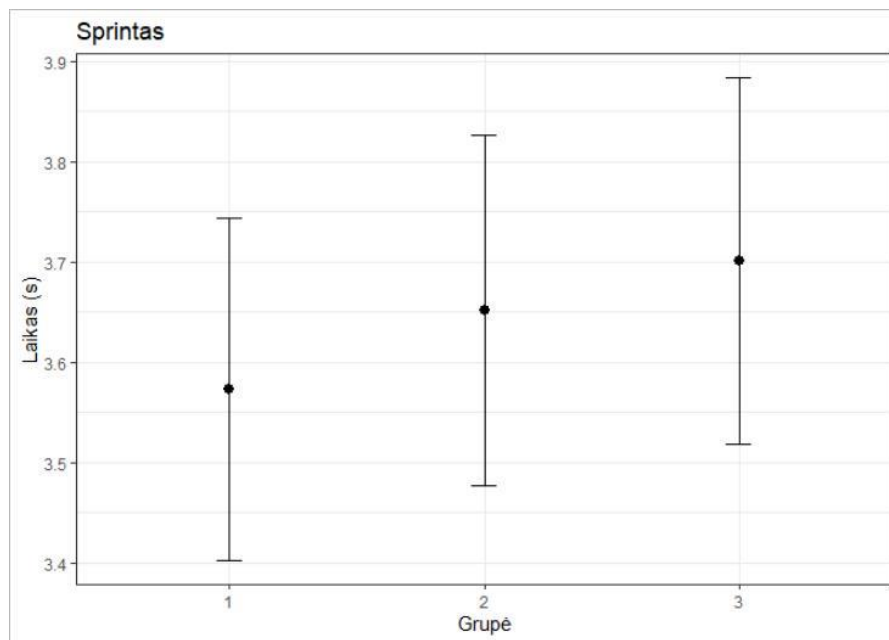
17 pav. Grupių lyginimo lentelė (R)

Pagal lentelę matome, kad ūgiai statistiškai reikšmingai skiriasi tarp visų grupių. Patikrinę R kvadrato reikšmę gauname, kad tik $< 0,01$ % masių lėmė skirtingi faktoriai (futbolo žaidimo lygiai), o > 99 % – individualūs žaidėjų skirtumai.

2.4.Sprintas



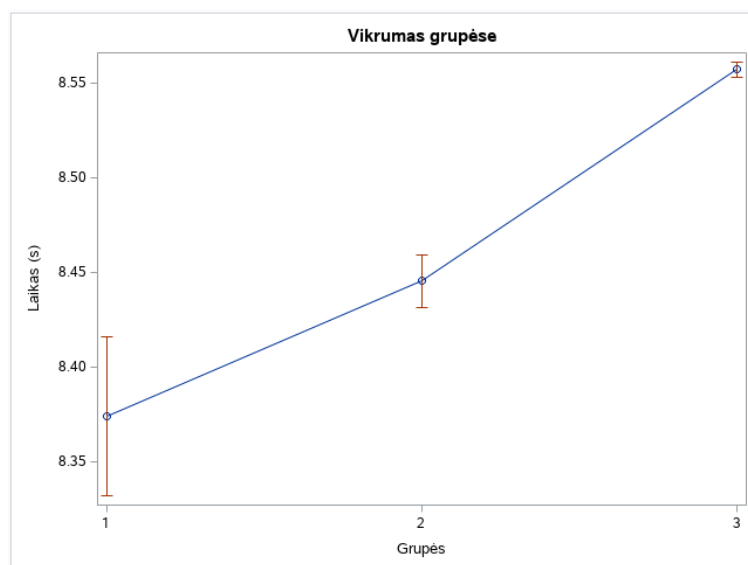
18 pav. Sprinto vidurkių ir standartinių paklaidų grafikas (SAS)



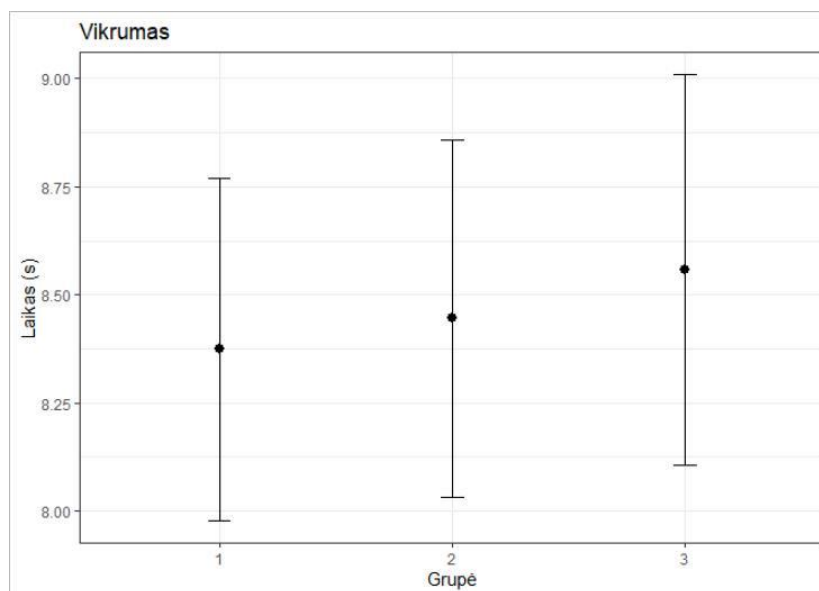
19 pav. Sprinto vidurkių ir standartinių nuokrypių grafikas (R)

20 metrų sprinto laikų vidurkiai didėja. Tikriname nulinę hipotezę, kad dispersijos yra lygios. Gauname, kad p reikšmė = 0,28 yra daugiau nei reikšmingumo lygmuo, todėl nulinės hipotezės neatmetame. Tikriname hipotezę apie vidurkių lygybę. Gauname, kad p reikšmė < 0,001, tai ji mažesnė už reikšmingumo lygmenį, todėl nulinę hipotezę atmetame. Sprinto laikai nėra vienodi. Patikrinus, kurios grupės skiriasi, gauname, kad visų grupių laikai yra tarpusavyje skirtingi. Patikrinę R kvadrato reikšmę gauname, kad tik < 0,01 % sprinto greičių lėmė skirtingi faktoriai (futbolo žaidimo lygiai), o > 99 % – individualūs žaidėjų skirtumai.

2.5. Vikrumas



20 pav. Vikrumo vidurkių ir standartinių paklaidų grafikas (SAS)



21 pav. Vikrumo vidurkių ir standartinių nuokrypių grafikas (R)

Tikriname nulinę hipotezę, kad dispersijos yra lygios.

Levene's Test for Homogeneity of Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
apl	2	1.1273	0.5636	5.88	0.0028
Error	13612	1304.4	0.0958		

22 pav. Dispersijų lyginimo lentelė (SAS)

```

Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group   2  8.7002 0.0001675 ***
      13612

```

23 pav. Dispersijų lyginimo lentelė (R)

Gauname, kad p reikšmė yra mažiau nei reikšmingumo lygmuo, todėl nulinę hipotezę atmetame, dispersijos tarp grupių nėra lygios. Todėl vidurkių palyginimui naudosime Welch ANOVA. Tikriname nulinę hipotezę, kad vidurkiai tarp grupių yra vienodi.

Welch's ANOVA for agi			
Source	DF	F Value	Pr > F
apl	2.0000	38.45	<.0001
Error	219.7		

24 pav. Vidurkių lyginimo lentelė (SAS)

One-way analysis of means (not assuming equal variances)

data: agi and apl
F = 38.447, num df = 2.00, denom df = 219.72, p-value = 4.819e-15

25 pav. Vidurkių lyginimo lentelė (R)

Gauname, kad p reikšmė mažesnė už reikšmingumo lygmenį, todėl nulinę hipotezę atmetame. Vikrumo testo laikai nėra vienodi. Tikriname vidurkių skirtumus tarp kiekvieno faktoriaus.

Comparisons significant at the 0.05 level are indicated by ***.				
apl Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
3 - 2	0.111908	0.075329	0.148486	***
3 - 1	0.183443	0.071425	0.295461	***
2 - 3	-0.111908	-0.148486	-0.075329	***
2 - 1	0.071535	-0.045556	0.188627	
1 - 3	-0.183443	-0.295461	-0.071425	***
1 - 2	-0.071535	-0.188627	0.045556	

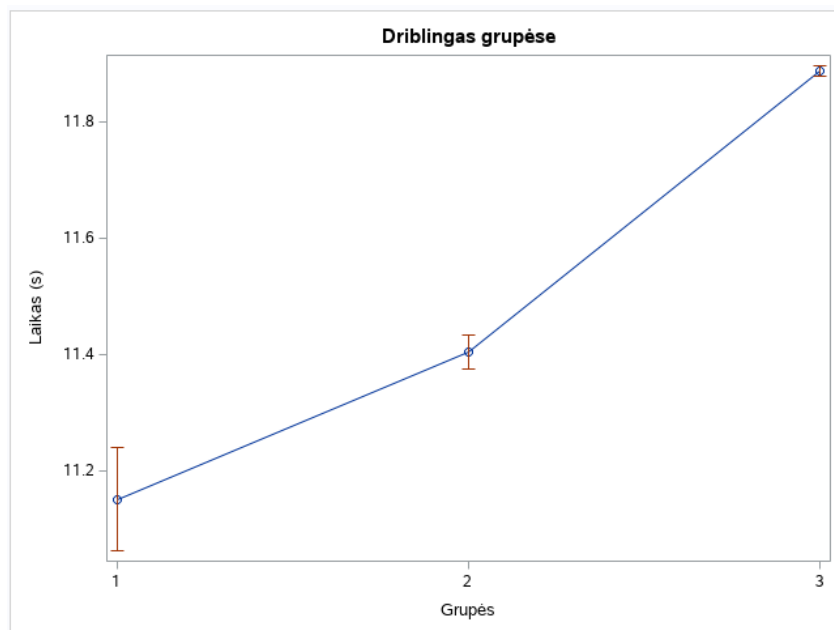
26 pav. Vidurkių skirtumo tarp faktorių lyginimo lentelė (SAS)

	1	2
2	0.45650	-
3	0.00037	2.3e-12

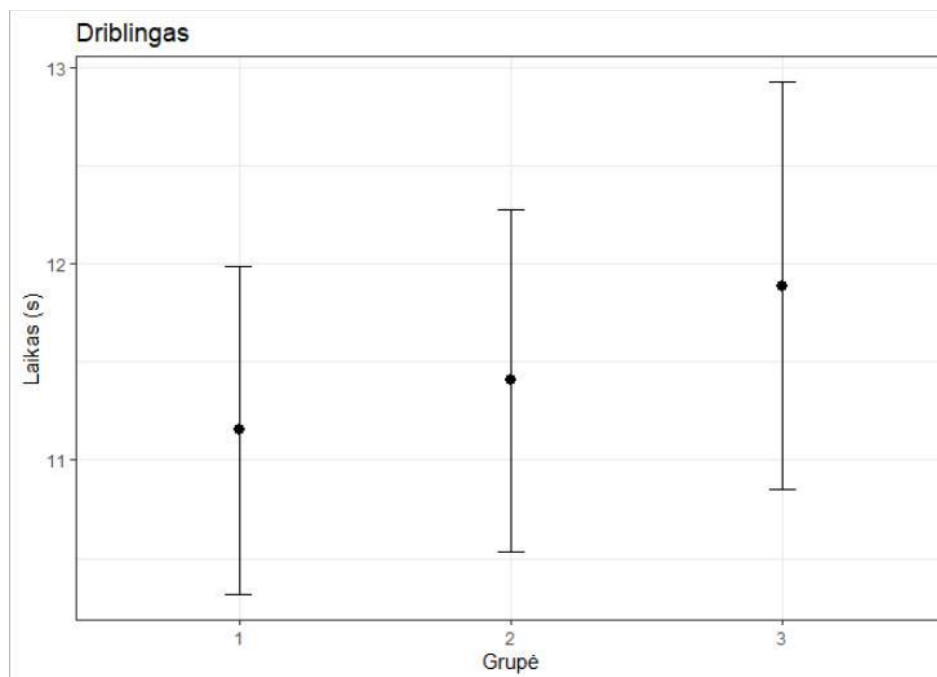
27 pav. Vidurkių skirtumo tarp faktorių lyginimo lentelė (R)

Gauname, kad pirmos (APL1) ir antros (APL2) grupės laikai statistiškai reikšmingai nesiskiria (nėra ***). O pirmos ir trečios (APL3) bei antros ir trečios laikai statistiškai reikšmingai skiriasi. Patikrinę R kvadrato reikšmę gauname, kad tik < 0,01 % vikrumo greičių lėmė skirtingi faktoriai (futbolo žaidimo lygiai), o > 99 % – individualūs žaidėjų skirtumai.

2.6.Driblingas



28 pav. Driblingo vidurkių ir standartinių paklaidų grafikas (SAS)



29 pav. Driblingo vidurkių ir standartinių nuokrypių grafikas (R)

Tikriname nulinę hipotezę, kad dispersijos yra lygios. Gauname, kad p reikšmė $< 0,001$, o tai yra mažiau nei reikšmingumo lygmuo, todėl nulinę hipotezę atmetame, dispersijos tarp grupių nėra lygios. Todėl vidurkių palyginimui naudosime Welch ANOVA. Tikriname nulinę hipotezę, kad vidurkiai tarp grupių yra vienodi. Gauname, kad p reikšmė $< 0,001$, o tai yra mažiau už reikšmingumo lygmenį, todėl

nulinę hipotezę atmetame. Driblingo testo laikai nėra vienodi. Tikriname vidurkių skirtumus tarp kiekvieno faktoriaus.

Comparisons significant at the 0.05 level are indicated by ***.				
apl Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
3 - 2	0.48325	0.39980	0.56670	***
3 - 1	0.73713	0.48157	0.99269	***
2 - 3	-0.48325	-0.56670	-0.39980	***
2 - 1	0.25388	-0.01325	0.52101	
1 - 3	-0.73713	-0.99269	-0.48157	***
1 - 2	-0.25388	-0.52101	0.01325	

30 pav. Vidurkių skirtumo tarp faktorių lyginimo lentelė (SAS)

```

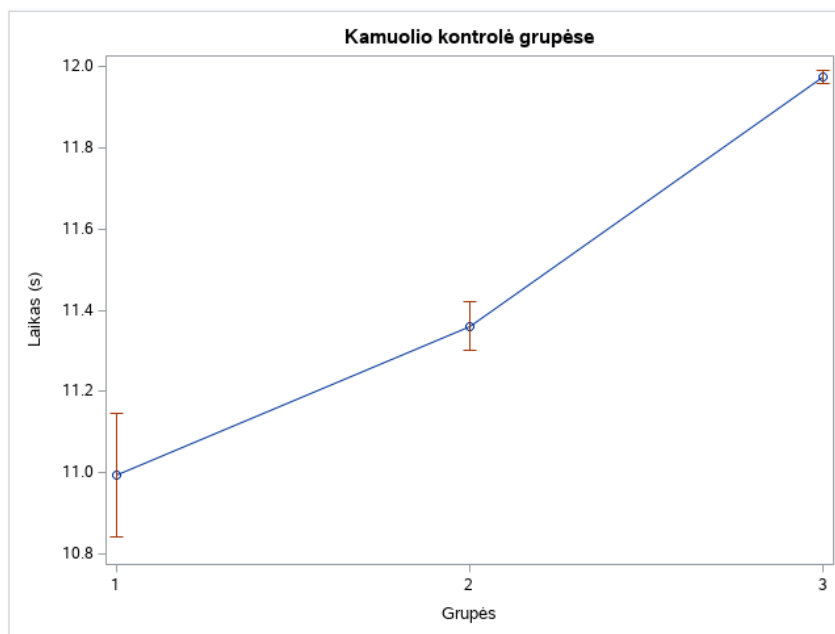
1      2
2 0.078 -
3 4.3e-11 < 2e-16

```

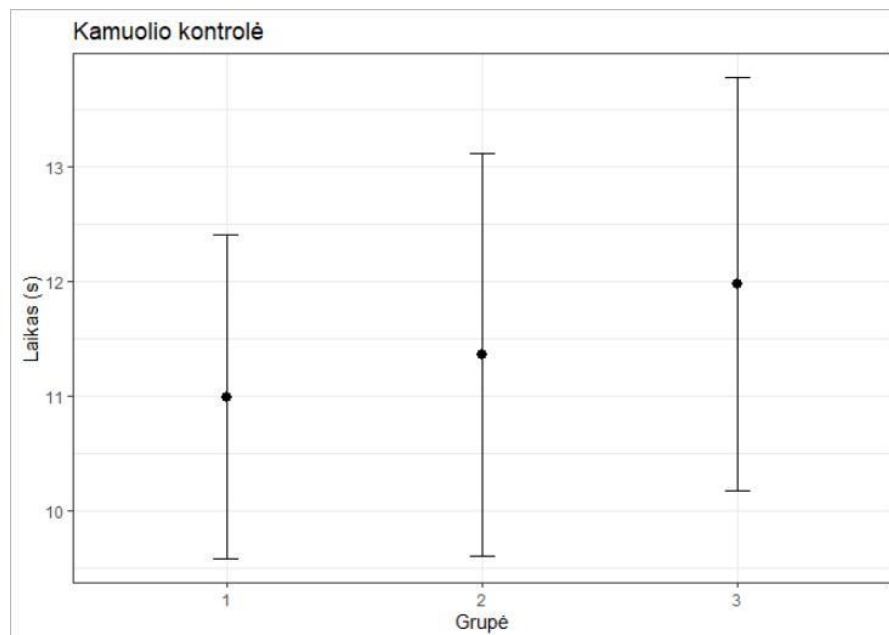
31 pav. Vidurkių skirtumo tarp faktorių lyginimo lentelė (R)

Gauname, kad APL1 ir APL2 grupių laikai statistiškai reikšmingai nesiskiria. O APL1 ir APL3 bei APL2 ir APL3 laikai statistiškai reikšmingai skiriasi. Patikrinę R kvadrato reikšmę gauname, kad tik 2 % driblingo greičių lėmė skirtingi faktoriai (futbolo žaidimo lygiai), o 98 % – individualūs žaidėjų skirtumai.

2.7. Kamuolio kontrolė



32 pav. Kamuolio kontrolės vidurkių ir standartinių paklaidų grafikas (SAS)



33 pav. Kamuolio kontrolės vidurkių ir standartinių nuokrypių grafikas (R)

Tikriname nulinę hipotezę, kad kamuolio kontrolės laiko dispersijos yra lygios. Gauname, kad p reikšmė = 0,16 yra daugiau nei reikšmingumo lygmuo, todėl nulinės hipotezės neatmetame. Tikriname hipotezę apie vidurkių lygybę. Gauname, kad p reikšmė < 0,001, tai ji mažesnė už reikšmingumo lygmenį, todėl nulinę hipotezę atmetame. Kamuolio kontrolės laikai nėra vienodi. Tikriname, kurių grupių laikai skiriasi tarpusavyje.

Comparisons significant at the 0.05 level are indicated by ***.				
apl Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
3 - 2	0.61277	0.46463	0.76091	***
3 - 1	0.97984	0.52708	1.43260	***
2 - 3	-0.61277	-0.76091	-0.46463	***
2 - 1	0.36706	-0.10626	0.84039	
1 - 3	-0.97984	-1.43260	-0.52708	***
1 - 2	-0.36706	-0.84039	0.10626	

34 pav. Vidurkių skirtumo tarp faktorių lyginimo lentelė (SAS)

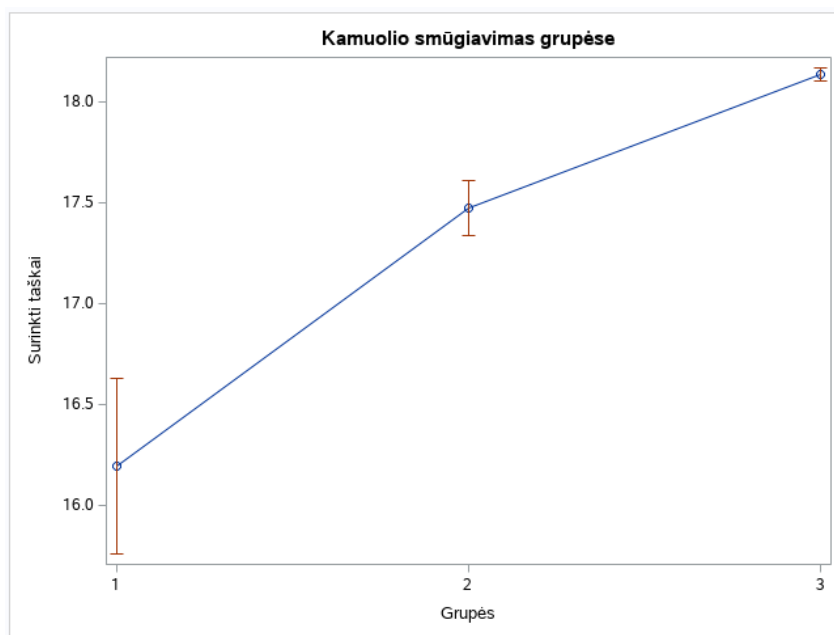
	1	2
2	0.21	-
3	1.2e-06	< 2e-16

35 pav. Vidurkių skirtumo tarp faktorių lyginimo lentelė (R)

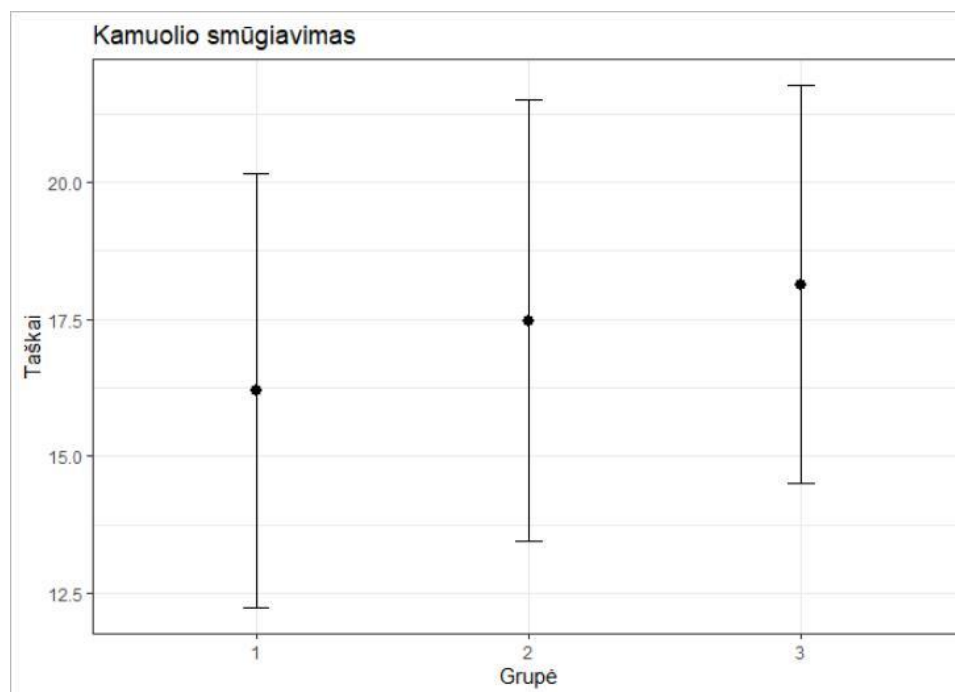
Gauname, kad APL1 ir APL2 grupių laikai statistiškai reikšmingai nesiskiria. O APL1 ir APL3 bei APL2 ir APL3 laikai statistiškai reikšmingai skiriasi. Patikrinę R kvadrato reikšmę gauname, kad tik

$< 0,01$ % kamuolio kontrolės greičio lėmė skirtingi faktoriai (futbolo žaidimo lygiai), o > 99 % – individualūs žaidėjų skirtumai.

2.8.Kamuolio smūgiavimas



36 pav. Kamuolio smūgiavimo vidurkių ir standartinių paklaidų grafikas (SAS)



37 pav. Kamuolio smūgiavimo vidurkių ir standartinių nuokrypių grafikas (R)

Tikriname nulinę hipotezę, kad kamuolio smūgiavimo taškų dispersijos yra lygios. Gauname, kad p reikšmė $< 0,001$, o tai yra mažiau nei reikšmingumo lygmuo, todėl nulinę hipotezę atmetame, dispersijos tarp grupių nėra lygios. Todėl vidurkių palyginimui naudosime Welch ANOVA. Tikriname nulinę hipotezę, kad vidurkiai tarp grupių yra vienodi. Gauname, kad p reikšmė $< 0,001$, o tai yra mažiau už reikšmingumo lygmenį, todėl nulinę hipotezę atmetame. Kamuolio smūgiavimo surinkti taškai nėra vienodi. Patikrinus, kurios grupės skiriasi tarpusavyje gauname, kad visų grupių laikai yra skirtingi. Patikrinę R kvadrato reikšmę gauname, kad tik $< 0,01$ % kamuolio smūgiavimo taškus lėmė skirtingi faktoriai (futbolo žaidimo lygiai), o > 99 % – individualūs žaidėjų skirtumai

2.9.Python programavimo kalba

Taip pat atlikome papildomą užduotį su Python kalba, tačiau išvados identiškos, dėl to apraše rezultatai nepateikiami.

IŠVADOS

Mūsų atlikto tyrimo ir tyrinėto straipsnio rezultatai sutapo identišškai. Gavome, kad tiek kūno sudėjimo, tiek žaidėjų įgūdžių parametrai statistiškai reikšmingai skyrėsi tarp žaidėjų profesionalumo lygio. Mažiausiai panašumų buvo tarp pirmosios ir trečiosios bei antrosios ir trečiosios grupės futbolininkų, kas leidžia teigti, kad futbolininkai, žaidžiantys top 5 Vokietijos lygose turi įgūdžių pranašumų prieš likusius neprofesionalių žaidėjų. Tokie požymiai kaip vikrumas, driblingas, kamuolio kontrolė statistiškai reikšmingai nesiskyrė tarp APL1 ir APL2 grupių, tad galime daryti išvadą, kad nebūtinai visi futbolininkai, žaidžiantys top 3 aukščiausiose lygose, pralenkia 4 ir 5 lygos žaidėjus.

ŠALTINIAI

- [1] „PLOS ONE“ tinklapis. Prieiga per internetą:
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0182211>