

Viimase 10 jalgpalli maailmameistrivõistluste analüüs

Aleksander Tamme

1/31/2019

1 Sissejuhatus

Töö eesmärgiks on uurida andmestikku, mis käsitleb jalgpalli maailmameistrivõistluste mängu, mis toimusid turniiridel vahemikus 1930-2014, andmestikust on puudu kõige viimane turniir, mis toimus aastal 2018. Uurimise alla võetakse andmestiku viimased 10 turniiri ehk aastad 1974-2014. Püstitame kolm hüpoteesi:

1. Euroopa meeskonnad on edukamad ehk enim mängu pidanud meeskondade top 15 domineerivad Euroopa meeskonnad.
2. Publikuarv mõjutab mängus löödud väravate hulka, mida suurem publikuarv, seda rohkem väravaid.
3. Teatavasti maailmameistrivõistlustel koduväljakueelist ei ole, sest turniir toimub iga kord ühes kindlas riigis(mõni turniir kahes riigis) ja igas mängus on meeskondade toetajate vahel pileteid jagatud võrdselt. Siiski on statistika mõttes vajalik igas mängus eraldada kodumeeskond ja võõrsilmeeskond. Väidame, et kodumeeskonnaks määratud meeskondadel on suurem võiduprotsent ja nad löövad rohkem väravaid kui võõrsilmeeskonnaks määratud meeskonnad.
4. Kõige suurema publikuhulgaga mängud, mis MM-turniiridel on olnud, jäid eelmisesse sajandisse, sest praegu ei saa turvalisuse kaalutlusel nii palju inimesi staadionile lubada.

2 Andmestik

2.1 Andmestiku kirjeldus

Andmestikus on 4572 rida ja 20 veergu, igal real on ühe konkreetse mängu info. Enne analüüsi peame andmestiku korrastama, hetkel on suur hulk ebavajalikke andmeid. Mängu kohta on 20 erinevat tunnust:

- aasta(kuna mäng toimus)
- kuupäev(mis kuupäeval ja kellaajal mäng toimus)
- turniirifaas(kas mäng leidis aset grupiturniiris või mõnes väljalangemisvoorus)
- staadion
- linn, kus staadion asub
- kodumeeskond
- kodumeeskonna löödud väravate arv
- võõrsilmeeskonna löödud väravate arv
- võõrsilmeeskond
- võidutingimus(kui mängul oli mõni lisatingimus, näiteks mäng otsustati penaltiseeriaga)
- publikuarv
- poolajaks löödud kodumeeskonna väravate arv
- poolajaks löödud võõrsilmeeskonna väravate arv
- kohtunik
- esimene abikohtunik
- teine abikohtunik
- vooru ID
- mängu ID
- kodumeeskonna initsiaalid
- võõrsilmeeskonna initsiaalid

2.2 Andmestiku korrastamine

Viime esmalt andmed sobivale kujule, vaja on eemaldada tühjad read, need turniirid, mida me ei analüüsi, samuti tuleb kustutada topeltkirjed, eemaldame ka veerud, mida me analüüsis ei uuri. Valime sellised read, kus on mängud turniiridelt 1974-2014. Kasutame meetodit `unique()`, et eemaldada tabeli ülevaatamisel märgatud 2014. aasta topeltkirjed. Jätame valimata need veerud, mida meil vaja ei lähe ehk kuupäev, võidutingimus, poolajaks löödud kodumeeskonna väravate arv, poolajaks löödud võõrsilmeeskonna väravate arv, kohtunik, esimene abikohtunik, teine abikohtunik, voo ID, mängu ID, kodumeeskonna initsiaalid, võõrsilmeeskonna initsiaalid. Tekitame juurde kaks veergu, kus on vastavalt mängu tulemus ja mängu väravate arv. Kuivõrd valisime ainult osade turniiride mängud, siis paneme ka reanumbrid ühest alates jooksma.

Peale korrastamist saime andmestiku, kus on 604 rida ja 9 veergu.

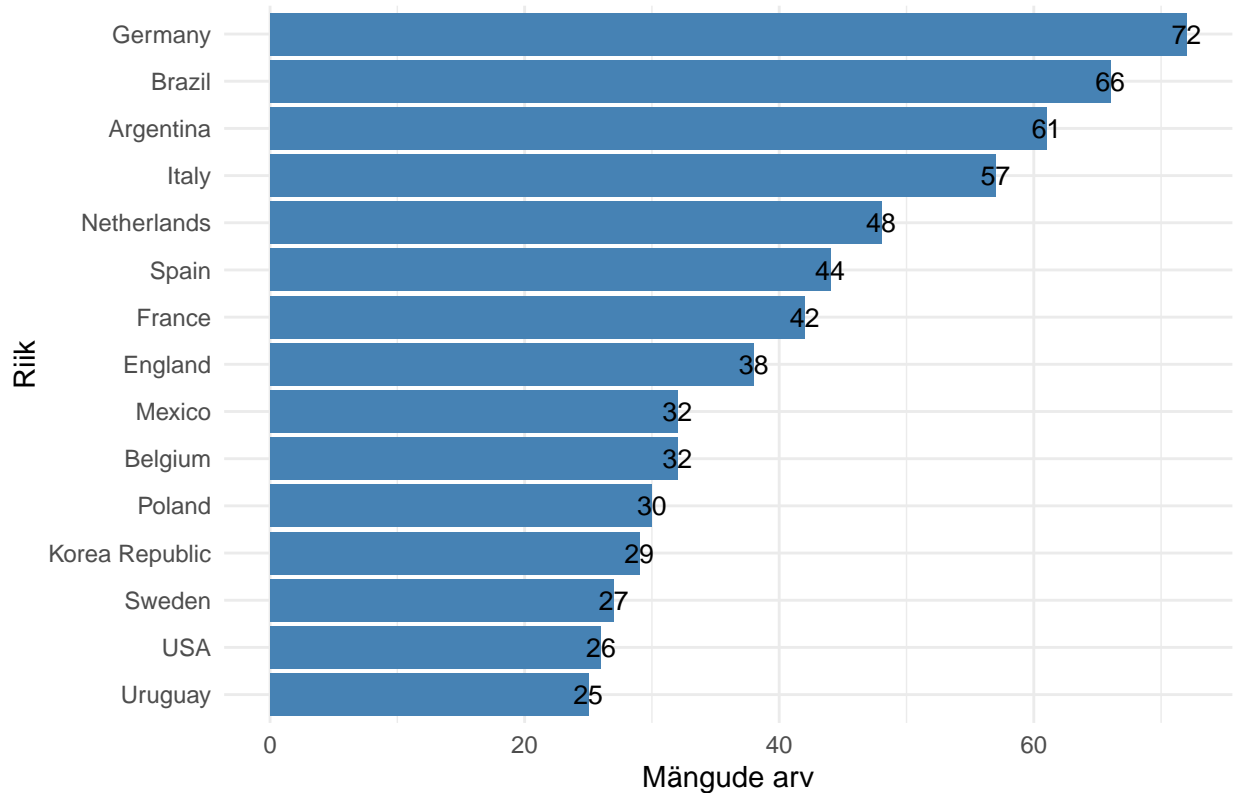
3 Analüüs

Välisel vaatlusel on näha, et osade meeskondade nimed on kuvatud valesti, esineb tähe- ja sümbolivigu. Lisaks tuleb Nõukogude Liit muuta Venemaaks ja Lääne-Saksamaa Saksamaaks.

Asume nüüd esimese hüpoteesi juurde ja selleks leiame kõikide turniiridel osalenud meeskondade mängitud mängude arvu. Selleks loeme kokku kõik sellised mängud, kus meeskond märgiti kodumeeskonnaks ja kõik sellised mängud, kus meeskond märgiti võõrsilmeeskonnaks ning liidame need kaks arvu. Kui juhtus, et mõni meeskond mängis kas ainult kodumeeskonnana või võõrsilmeeskonnana, siis esineb tabelis NA väärtusi. Selle ennetamiseks asendame tabelis kõik NA väärtused nullidega.

Leiame jalgpalli maailmameistrivõistlustel enim mängu pidanud riikide top 15. Selleks kasutame väikest vahetabelit, mida näha ei ole, kuhu salvestame kõikide turniiridel osalenud meeskondade mängude arvu.

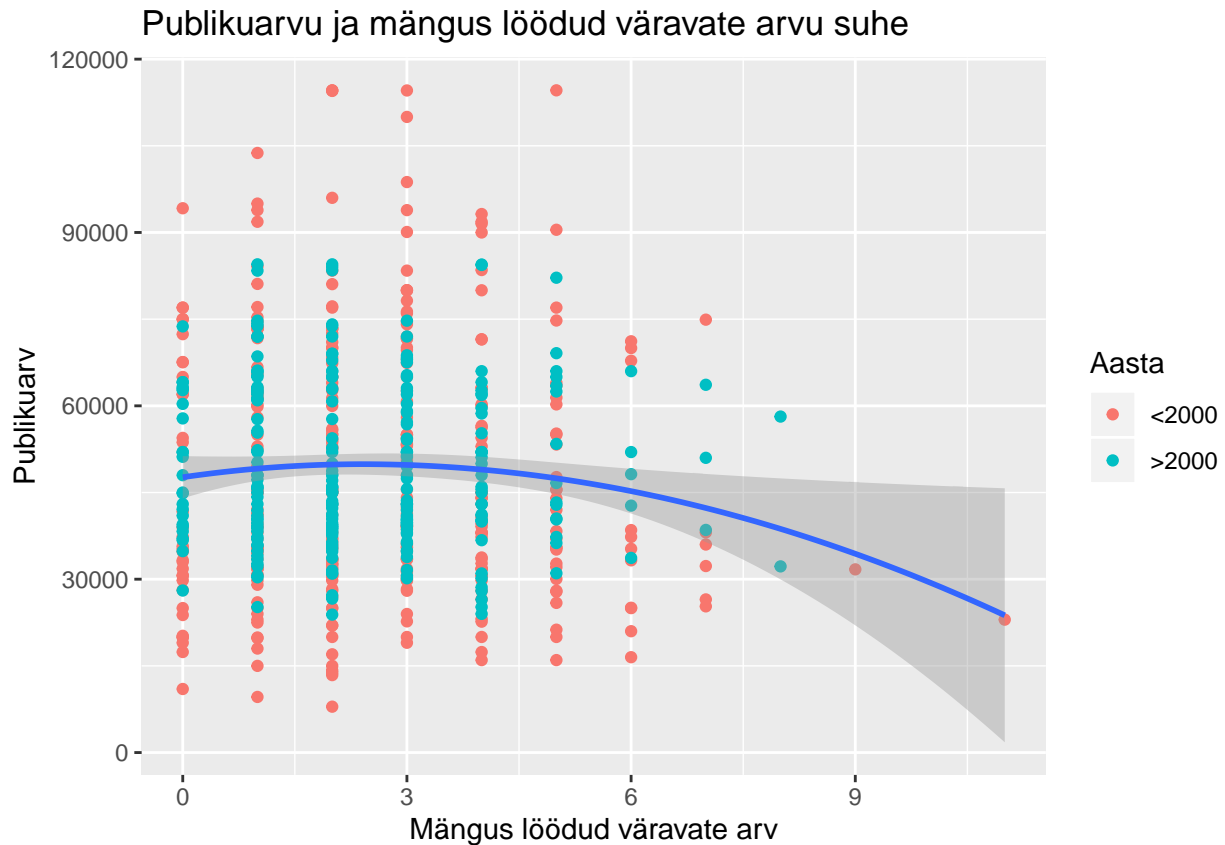
Enim mängu pidanud riikide top 15



15-st enim mängu pidanud meeskonnast üheksa on Euroopa meeskonnad, sealhulgas esimesest üheksast tiimist on seitse Euroopa tiimid. Enim mängu on maailmameistrivõistlustel pidanud Saksamaa, 72 mängu.

Püstitasime hüpoteesi, et Euroopa meeskonnad domineerivad enim mängu pidanud riikide top 15 ehk on olnud edukamad. Hüpotees osutus tõeseks.

Teise hüpoteesi uurimiseks koostame joonise, kus punktidenäha kõik mängud aastatel 1974-2014. Arusaamise lihtsustamiseks pole iga turniir eraldi värviga kirjas, turniirid on jagatud kas enne 2000. aastat toimunuks või peale 2000. aastat toimunuks.



Mingit suurt seaduspära jooniselt ei leia, publikuarvul ja mängus löödud väravate arvu ei tundu seost olevat. Suureskooriliste mängude puhul (5+ väravat) võib siiski märgata trendi, et selliseid mängu käib vaatamas keskmisest väiksem arv inimesi. Tõsi, selliseid mängu on olnud palju vähem, seega on info hulk väiksem. Suure skooriga mängud tulevad tihti kahe meeskonna suurest kvaliteedivahest. Võib teoretiseerida, et selliseid mängu, kus nõ võitja on ette teada, käib vaatamas vähem inimesi, kui kahe võrdse meeskonna mõõduvõttu. Jooniselt saab välja lugeda ka seda, et maailmameistrivõistluste suurimad publikunumbrid sündisid eelmisel sajandil. Ka on märgata, et suureskoorilised mängud on jäänud harvemaks kui varasematel MM-idel, mõne erandiga.

Püstitasime hüpoteesi, et suurem publikuarv mängul tähendab suuremat mängus löödud väravate hulka. Hüpoteesi ei saa kinnitada.

Nüüd vaatame meie kolmandat püstitatud hüpoteesi, mis ütleb, et kodumeeskonnaks määratud meeskonnad võivad tihedamini ja löövad rohkem väravaid. See on püstitatud küsimustest kõige põnevam, sest kui tavalistes jalgpallimängudes on koduväljakueelis reaalne uuritud asi, siis siin mingit eelist justkui ei tohiks olla, sest maailmameistrivõistluste mängud peetakse niikuinii neutraalsel pinnal.

Ometi näeme jubavõrreldes kodumeeskonna löödud väravate arvu keskmist ja võõrsilmeeskonna löödud väravate arvu keskmist, et mingi eelise kodumeeskonnaks määramine ikkagi annab. Kodumeeskond lööb mängus keskmiselt 1.46 väravat, võõrsilmeeskond seevastu ainult 1.07 väravat, koguni 0,39 väravat mängu kohta vähem. Järgmiseks leiame, mitu väravat on löödud kokku kodumeeskondade poolt ja mitu võõrsilmeeskondade poolt.

Kodumeeskonnad on löönud 57.7 % kõigist väravatest(881 väravat) maailmameistrivõistlustel ja võõrsilmeeskonnad 42.3 % väravatest(647 väravat)

Table 1: Tulemuste jaotumine

jaotumine	Freq
Kodumeeskonna võit	282
Viik	170
Võõrsilmeeskonna võit	152

Tulemuste jaotumine kodu- ja võõrsilmeeskonna vahel



Kokku peeti 1974-2014 MM turniiridel 604 mängu, nendest 282 ehk 46,7% lõppesid kodumeeskonna võiduga, 170 ehk 28,1% lõppesid viigiga ja 152 ehk 25,2% lõppes võõrsilmeeskonna võiduga. Siin tuleb kodumeeskonna eelis eriti selgelt välja, ligi pooled mängudest on võitnud kodumeeskond, samal ajal kui võõrsilmeeskonnad on suutnud ainult neljandiku peetud mängudest võita.

Püstitasime hüpoteesi, et kodumeeskonnaks määratud meeskond võidab rohkem mänge ja lööb rohkem väravaid kui võõrsilmeeskonnaks määratud meeskond. Hüpotees leidis kinnitust.

Table 2: 15 kõige suurema publikuarvuga mängu

	Aasta	Kodumeeskond	Kodumeeskonna väravad	Võõrsilmeeskonna väravad	Võõrsilmeeskond	Publikuarv
148	1986	Mexico	1	1	Paraguay	114600
180	1986	Argentina	3	2	Germany	114600
166	1986	Mexico	2	0	Bulgaria	114580
176	1986	Argentina	2	1	England	114580
178	1986	Argentina	2	0	Belgium	114500
136	1986	Belgium	1	2	Mexico	110000
158	1986	Iraq	0	1	Mexico	103763
171	1986	England	3	0	Paraguay	98728
129	1986	Bulgaria	1	1	Italy	96000
77	1982	Argentina	0	1	Belgium	95000
284	1994	Brazil	0	0	Italy	94194
247	1994	USA	2	1	Colombia	93869
257	1994	USA	0	1	Romania	93869
240	1994	Cameroon	2	2	Sweden	93194
237	1994	Colombia	1	3	Romania	91856

Näeme, et kõige suurema publikuarvuga mängud toimusid eelmisel sajandil nagu eeldatud. Esikohta jagavad Mehhiko ja Paraguai 1-1 viik 1986. aasta MM-il ja Argentiina 3-2 võit Saksamaa üle sama aasta turniiril. Neid mänge jälgis 114 600 inimest. Tänapäeval pole Euroopas praktiliselt ühtegi sellist staadioni, kuhu nii palju inimesi mahuks.

Püstitasime hüpoteesi, et kõige suurema publikuarvuga mängud jäid eelmisesse sajandisse. Hüpotees osutus tõeseks.

4 Kokkuvõte

Käesolev projekt uuris 1974-2014 jalgpalli maailmameistrivõistlusi. Andmestik sisaldas nendest turniiridest põhilist informatsiooni, nagu mängude tulemused jne. Esitati 4 hüpoteesi, millest 3 leidsid kinnitust. Püstitatud hüpoteesid:

1. Turniiridel on rohkem mänge pidanud Euroopa meeskonnad. Hüpotees osutus tõeseks.
2. Mida suurem publikuarv, seda rohkem lüüakse väravaid. Hüpoteesi ei saa kinnitada.
3. Kodumeeskonnaks määratud meeskond lööb rohkem väravaid ja omab suuremat võiduanssi. Hüpotees osutus tõeseks.
4. Kõige suurema publikuarvuga mängud jäid eelmisesse sajandisse. Hüpotees osutus tõeseks.