

Kursinis darbas. Universitetų reitingų palyginimas

Simas Vencius ir Edgaras Petrauskas, Ekonometrija, 3 kursas

2016 m. spalio 12 d.

Ivadas

Kiekvienais metais Lietuvoje ir pasaulyje daugybę diskusijų sukelia universitetų reitingai. Egzistuoja šimtai skirtingų nacionalinių ir tarptautinių reitingų sistemų ir sudėtinga suprasti, kodėl skirtingos sistemos kai kurias aukštojo mokslo įstaigas įvertina kardinaliai skirtingai. Visuomenėje ne be reikalo kyla diskusijos dėl universitetų reitingavimo pagrįstumo. Reitingavimas turi ir didelę ekonominę įtaką - geriau reitinguoti universitetai turi geresnes galimybes pritraukti gabesnius studentus ir išlaikyti aukštą mokslo lygį, sulaukti pripažinimo ir didesnių investicijų. Ekonomiškai stiprios šalys dažniausiai turės ir geriausius universitetus.

Šaltinyje <https://www.kaggle.com/mylesoneill/world-university-rankings> pateikiami trys kiekvienais metais sudaromi skirtingi tarptautiniai universitetų reitingai. **Times Higher Education World University Ranking** - įkurtas Jungtinėje Karalystėje 2010 metais, **Academic Ranking of World Universities** (dar žinomas kaip **Shanghai** reitingas) - įkurtas Kinijoje 2003 metais ir **Center for World University Rankings** (toliau - **Cwur**) - Saudo Arabijoje 2012 metais. Kiekvieno reitingo duomenys pateikti skirtingose lentelėse.

Kursiniame darbe nagrinėsime, kokie reitingų panašumai ir ar yra ryškių skirtumų tarp universitetų reitingų ir jei yra, kas juos lemia. Ar šalys daro įtaką reitingui ir ar didesnės išlaidos mokslui lemia geresnį universiteto reitingą.

Duomenų apžvalga

- “timesData.csv” pateikiami 2011-2016 metų duomenys iš Times reitingo:
 - world_rank - universiteto vieta reitinge. Yra intervalinių rangų ir lygių rangų (pvz. =94 ir 201-250).
 - university_name - universiteto pavadinimas.
 - country - šalis.
 - teaching - mokymo balas.
 - international - universiteto įvaizdžio tarptautiniame lygmenyje balas.
 - research - universiteto tyrimų balas.
 - citations - citavimo balas.
 - income - pajamų balas.
 - total_score - galutinis balas, naudojamas galutinio reitingo nustatymui.
 - num_students - studentų skaičius universitete.
 - student_staff_ratio - studentų skaičius padalintas iš darbuotojų skaičiaus.
 - international_students - tarptautinių studentų dalis procentais.
 - female_male_ratio - vyrų ir moterų santykis.
 - year - reitingavimo metai (2011-2016 m.).
-
- “shanghaiData.csv” pateikiami 2005-2015 metų duomenys iš Shanghai reitingo.
 - world_rank - universiteto vieta reitinge. Yra intervalinių rangų ir lygių rangų (pvz. 101-152).
 - university_name - universiteto pavadinimas.

- national_rank - universiteto reitingas savoje šalyje.
 - total_score - galutinis balas, naudojamas galutinio reitingo nustatymui.
 - alumni - alumni balas, kurį nulemia alumni, gavusių Nobelio premiją arba Fields medalį, skaičius.
 - award - premijos balas, kurį nulemia įstaigos darbuotojų, gavusių Nobelio premijas fizikos, chemijos, medicinos, ekonomikos ir Fields medalį matematikos srityse, skaičius.
 - hici - HiCi balas, kurį nulemia citavimų skaičius
 - ns - N&S balas, kurį nulemia Nature and Science žurnale įtrauktų publikacijų skaičius.
 - pub - PUB balas, kurį nulemia Science Citation Index-Expanded and Social Science Citation Index įtrauktų publikacijų skaičius.
 - pcp - PCP balas, kurį nulemia penkių viršutinių parametų svertiniai balai padalinti iš pilno etato akademikų skaičiaus.
 - year - reitingavimo metai (2005-2015 m.).
-

- “cwurData.csv” pateikiami 2012-2015 metų duomenys iš Cwur reitingo.

- world_rank - universiteto vieta reitinge.
 - university_name - universiteto pavadinimas.
 - country - šalis.
 - national_rank - universiteto reitingas savoje šalyje.
 - quality_of_education - švietimo kokybės balas.
 - alumni_employment - alumni darbo lygio balas.
 - quality_of_faculty - fakulteto kokybės balas.
 - publications - publikacijų balas.
 - influence - įtakos balas.
 - citations - citavimo balas.
 - broad_impact - h-index įtakos balas (tik 2014 ir 2015 m. duomenys).
 - patents - patentų balas.
 - score - galutinis balas, naudojamas galutinio reitingo nustatymui.
 - year - reitingavimo metai (2012-2015).
-

Reitingų sandaros grafinė apžvalga

```
par(mfrow=c(2,2))
pie_cwur <- c(25, 25, 25, 25)
pie(pie_cwur, main="Cwur reitingo sandara", col=rainbow(length(pie_cwur)),
    labels=c("Quality of Education", "Alumni Employment", "Quality of Faculty", "Kita"))

pie_shanghai <- c(10, 20, 20, 20, 20, 10)
pie(pie_shanghai, main="Shanghai reitingo sandara", col=rainbow(length(pie_shanghai)),
    labels=c("Alumni", "Award", "HiCi", "N&S", "PUB", "PCP"))

pie_times <- c(30, 30, 30, 10)
pie(pie_times, main="Times reitingo sandara", col=rainbow(length(pie_times)),
    labels=c("Teaching", "Research", "Citations", "Kita"))
```

Cwur reitingo sandara



Shanghai reitingo sandara



Times reitingo sandara



Iš visų trijų reitingų imsime pirmus šimtukus, nes po pirmo šimtuko Shanghai reitinge prasideda intervalinis rangavimas. Taip pat ir Times reitinge nuo 200-osios vietos prasideda intervalinis reitingavimas, todėl negalima išsibrėžti grafikų, kad matytųsi priklausomybė tarp reitingo ir tam tikro parametro.

Times, Shanghai ir Cwur reitingai atitinkamai (2015 m.):

	world_rank	university_name	country	teaching	international	research
1403	1	California Institute of Technology	United States of America	92.2	67.0	98.0
1404	2	Harvard University	United States of America	92.9	67.6	98.0
1405	3	University of Oxford	United Kingdom	88.6	90.7	97.0
1406	4	Stanford University	United States of America	91.5	69.0	96.0
1407	5	University of Cambridge	United Kingdom	89.7	87.8	95.0
1408	6	Massachusetts Institute of Technology	United States of America	89.1	84.3	88.0

	world_rank	university_name	national_rank	total_score	alumni	award	hic
4398	1	Harvard University	1	100.0	100.0	100.0	100.0
4399	2	Stanford University	2	73.3	40.7	89.6	80.1
4400	3	Massachusetts Institute of Technology (MIT)	3	70.4	68.2	80.7	60.6
4401	4	University of California, Berkeley	4	69.6	65.1	79.4	66.1
4402	5	University of Cambridge	1	68.8	77.1	96.6	50.8
4403	6	Princeton University	5	61.0	53.3	93.4	57.1

	world_rank	institution	country	national_rank	quality_of_education
1201	1	Harvard University	USA	1	1
1202	2	Stanford University	USA	2	9
1203	3	Massachusetts Institute of Technology	USA	3	3
1204	4	University of Cambridge	United Kingdom	1	2
1205	5	University of Oxford	United Kingdom	2	7
1206	6	Columbia University	USA	4	13

Pažymėtina, jog reitingai parametrus vertina skirtingais metodais. Times ir Shanghai parametrai įvertinti šimtabalėje skalėje, o Cwur reitinge kiekvienas parametras yra suranguotas.

Kadangi reitingo Cwur parametrai yra suranguoti, tas pats daroma su Times ir Shanghai reitingų citavimo parametrais, kad būtų galima palyginti grafikus. Rangavimo metodas “first”, nes taip jau yra suranguoti Cwur reitingo parametrų duomenys.

```
times_cit_rank <- rank((-times2015$citations), ties.method="first")
shanghai_cit_rank <- rank((-shanghai2015$hci), ties.method="first")

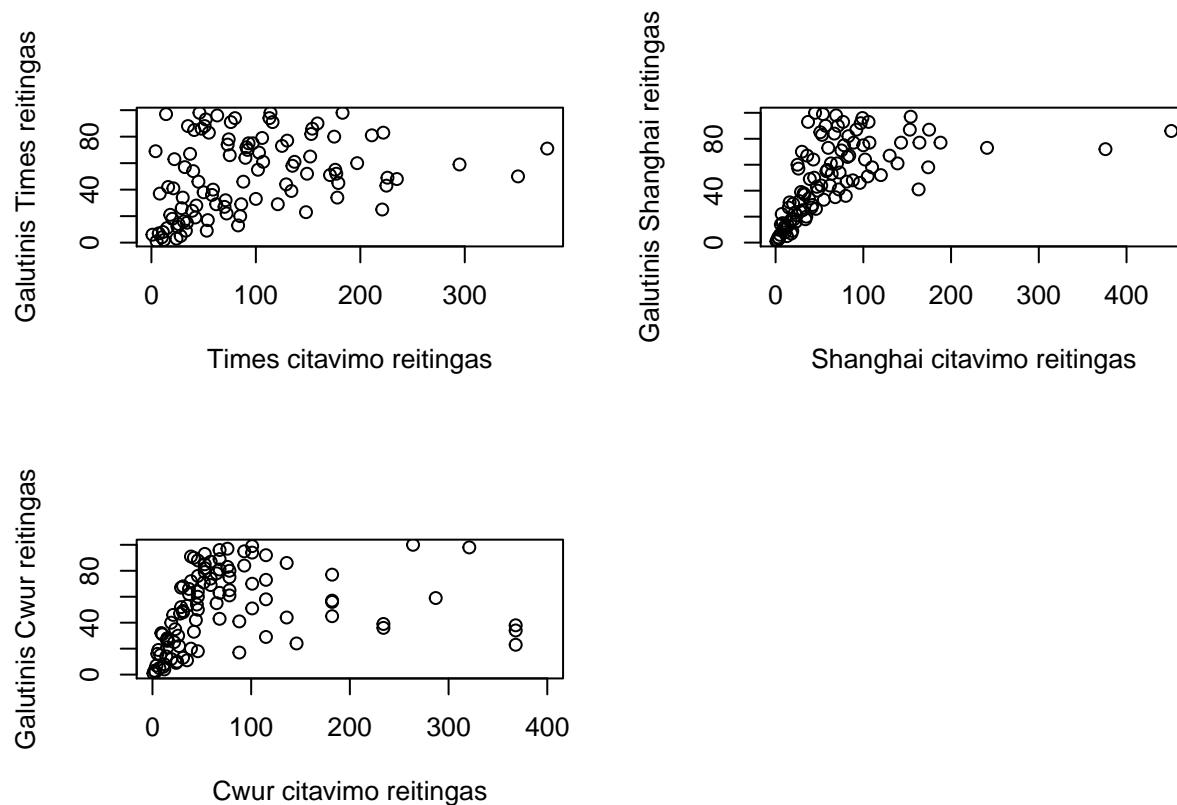
# Imame pirmą šimtuką

times_cit_rank100 <- times_cit_rank[1:100]
shanghai_cit_rank100 <- shanghai_cit_rank[1:100]

# Paverčiam į numeric, kad būtų galima skaičiuoti koreliaciją

sub_times$world_rank <- as.numeric(as.character(sub_times$world_rank))
sub_shanghai$world_rank <- as.numeric(as.character(sub_shanghai$world_rank))

par(mfrow=c(2,2))
plot(times_cit_rank100, sub_times$world_rank, ylab="Galutinis Times reitingas", xlab="Times citavimo rei
plot(shanghai_cit_rank100, sub_shanghai$world_rank, ylab="Galutinis Shanghai reitingas", xlab="Shanghai
plot(sub_cwur$citations, sub_cwur$world_rank, xlim=c(0,400), ylab="Galutinis Cwur reitingas", xlab="Cwur
```



Stipriausia priklausomybė atrodo Shanghai reitinge. Žiūrime koreliacijas:

```
cor(shanghai_cit_rank100, sub_shanghai$world_rank)
```

```
## [1] 0.540448
```

```
cor(sub_cwur$citations, sub_cwur$world_rank)
```

```
## [1] 0.176007
```

```
cor(times_cit_rank100, sub_times$world_rank)
```

```
## [1] 0.3204428
```

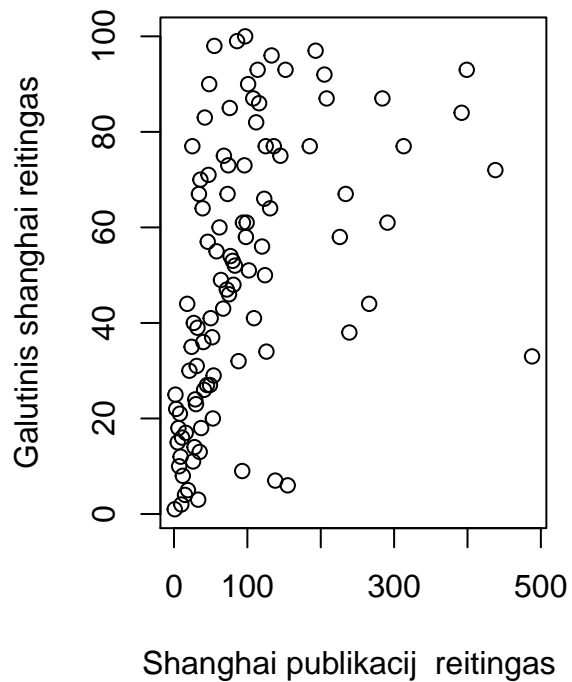
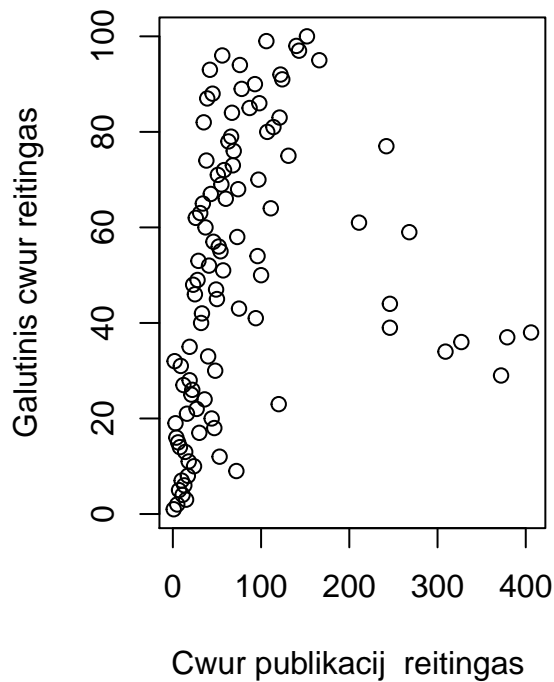
Koreliacija tarp citavimo reitingo ir galutinio reitingo stipriausia Shanghai reitinge.

Dabar tikrinsime priklausomybę tarp publikacijų reitingo ir galutinio reitingo. Ranguojame Shanghai reitingo universitetus pagal publikacijų parametą ir paimame pirmą šimtuką. Times reitingas publikacijų kaip atskiro kriterijaus nepateikė, todėl pagal publikacijas lyginsime tik Shanghai ir Cwur reitingus.

```
shanghai_pub_rank <- rank((-shanghai2015$pub), ties.method=c("first"))
shanghai_pub_rank100 <- shanghai_pub_rank[1:100]
```

```
par(mfrow=c(1,2))
```

```
plot(sub_cwur$publications, sub_cwur$world_rank, ylab="Galutinis cwur reitingas", xlab="Cwur publikacijų reitingas")
plot(shanghai_pub_rank100, sub_shanghai$world_rank, ylab="Galutinis shanghai reitingas", xlab="Shanghai publikacijų reitingas")
```



```
# Sunku pasakyti, kuriame reitinge didesnė priklausomybė tarp publikacijų reitingo ir galutinio reitin
cor(sub_cwur$publications, sub_cwur$world_rank)
```

```
## [1] 0.2275657
```

```
cor(shanghai_pub_rank100, sub_shanghai$world_rank)
```

```
## [1] 0.4356668
```

```
# Didesnė priklausomybė Shanghai reitinge
```

Imame Shanghai reitingo pirmus 100 universitetų ir išrenkame tuos pačius universitetus iš Times reitingo bei atliekame Spearmano testą galutinių reitingų priklausomybei nustatyti. Testo hipotezė:

H0: kintamieji nekoreliuoja

H1: kintamieji koreliuoja

Imame tik pirmą 100-tuką, nes kiekviename reitinge yra skirtingas skaičius universitetų, todėl jei imsime didesnį jų skaičių, galime kai kurių universitetų iš Shanghai reitingo nerasti Times reitinge ir atvirkščiai.

```
sub_times1 <- times2015[,c("university_name", "world_rank")]
sub_shanghai1 <- shanghai2015[c(1:100),c("university_name", "world_rank")]
# Eilutėms priskiriami vardai pagal universiteto pavadinimą
time_names_vector <- as.vector(sub_times1$university_name)
```

```

row.names(sub_times1) <- time_names_vector
shanghai_names_vector <- as.vector(sub_shanghai1$university_name)
row.names(sub_shanghai1) <- shanghai_names_vector
# Sujungiama į bendrą lentelę
bendra <- merge(sub_shanghai1, sub_times1, by = "row.names", all = FALSE)
bendra1 <- bendra[match(row.names(sub_shanghai1), bendra[, "Row.names"]), -1]
bendra2 <- na.omit(bendra1)
bendra2 <- bendra2[, c("world_rank.x", "world_rank.y")]
colnames(bendra2) <- c("shanghai_world_rank", "times_world_rank")
# Universitetui, kuris pateko į reitingo intervalą 226-250 priskiriamas reitingas 238, o kuris pateko į
bendra2[which(bendra2$times_world_rank=="226-250"), "times_world_rank"] = 238
bendra2[which(bendra2$times_world_rank=="201-225"), "times_world_rank"] = 213

# Paverčiame numeric, kad būtų galima atlikti testą
bendra2$times_world_rank <- as.numeric(as.vector(bendra2$times_world_rank))
bendra2$shanghai_world_rank <- as.numeric(as.vector(bendra2$shanghai_world_rank))
cor.test(bendra2$times_world_rank, bendra2$shanghai_world_rank, method = "spearman")

```

```

## Warning in cor.test.default(bendra2$times_world_rank,
## bendra2$shanghai_world_rank, : Cannot compute exact p-value with ties

```

```

##
## Spearman's rank correlation rho
##
## data: bendra2$times_world_rank and bendra2$shanghai_world_rank
## S = 10345, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.80256

```

H0 atmetame, kintamieji statistiškai reikšmingai koreliuoja. Vadinas Shanghai reitingo pirmojo šimtuko universitetai koreliuoja su Time reitingo tais pačiais universitetais.

Priklausomybė tarp reitingų matoma ir grafiškai.

```

plot(bendra2$times_world_rank, bendra2$shanghai_world_rank, xlab="Galutinis Times reitingas", ylab="Galutinis Shanghai reitingas")

```

