Izlases dizaina optimizācija problemātika un daži rezultāti

Mārtiņš Liberts

Latvijas Universitāte
LR Centrālā statistikas pārvalde

23.02.2012.

Saturs

Populācija

- Galīga populācija ar apjomu N

$$U = \{1,2, ..., N\}$$

- Populācija ir elementu kopa

$$k \in U$$

- Katram populācijas elementam ir piekārtotas vairākas pazīmes

$$y_k, z_k, \ldots$$

Populācija

- Populācijas parametri

$$Y = \sum_U y_k$$

$$\overline{Y} = \frac{1}{N} \sum_U y_k$$

$$R = \frac{Y}{\overline{Z}} = \frac{\overline{Y}}{\overline{\overline{Z}}}$$

Izlase

- Izlase s ar apjomu n tiek atlasīta no populācijas U ar zināmu varbūtību $p(s)$
- Izlasē iekļaušanas varbūtības katram elementam $k \in U$ ir

$$\pi_k = \sum_{s \ni k} p(s) = Pr(k \in s)$$

- Populācijas parametram var konstruēt novērtētāju $\hat{\theta}$, izmantojot s

$$\hat{\theta} = \hat{\theta}(s)$$

Izmaksas

- Apsekojuma organizācijas izmaksas

$$c_T = c_F + c_V$$

- Pieņēmums: mainīgās izmaksas ar atkarīgas no s

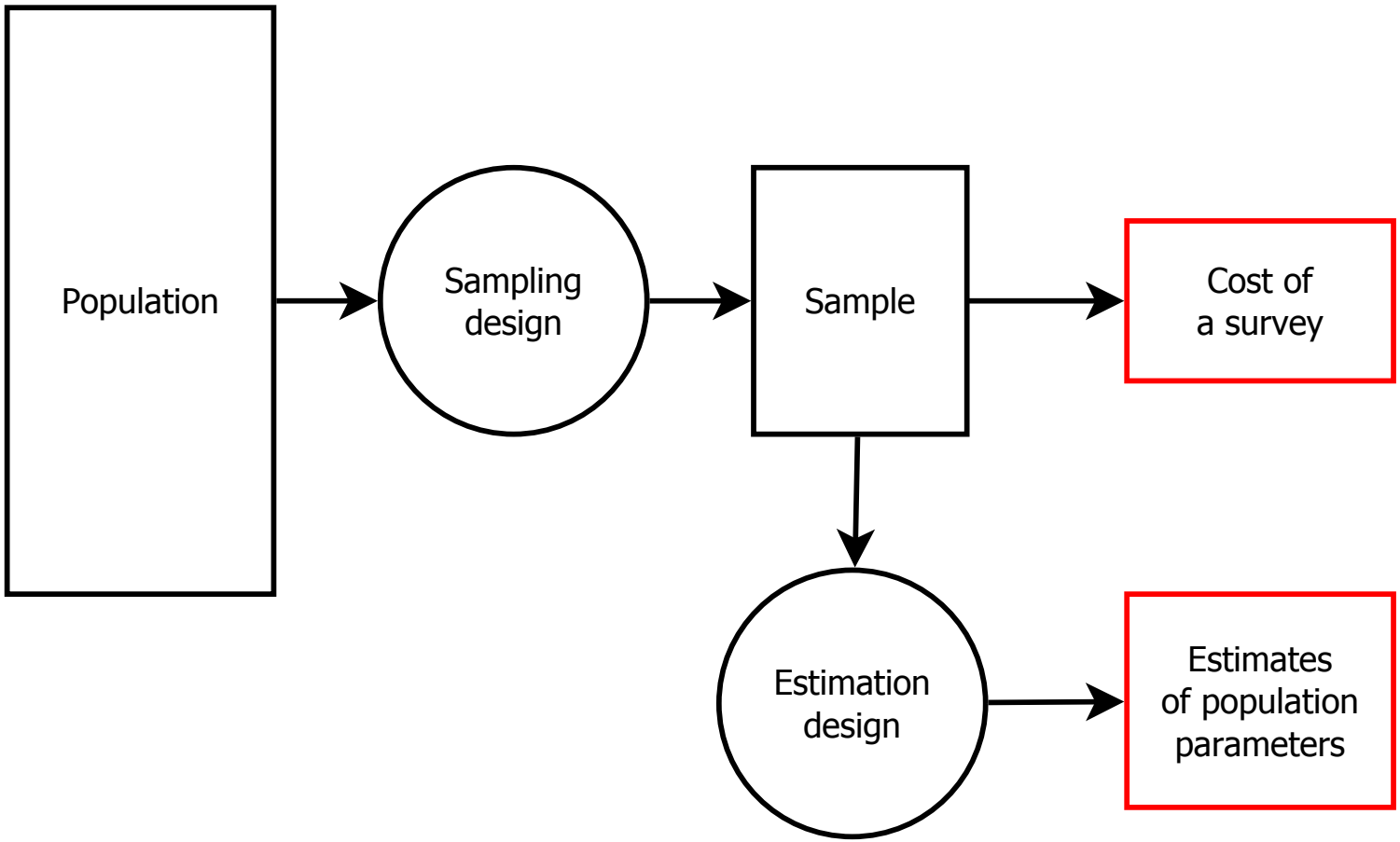
$$c_V = c_V(s)$$

- Piemēram:
  - Mainīgās izmaksas ir atkarīgas no $|s| = n$
  - Mainīgās izmaksas ir atkarīgas no izlases vienību ģeogrāfiskā izvietojuma, ja apsekojums tiek veikts, izmantojot personīgās intervijas (PAPI vai CAPI)

Dizaina optimizācija

- $\hat{\theta}$ un c_V ir gadījuma lielumi, jo izlase s tiek atlasīta kā varbūtiskā izlase
- Varam pētīt $E\left(\hat{\theta}\right)$, $V\left(\hat{\theta}\right)$, $E(c_V)$, $V(c_V)$ īpašības
- Praksē ir vēlams sasniegt vairākus mērķus:
  - $E\left(\hat{\theta}\right) \to \theta$
  - $V\left(\hat{\theta}\right) \to \min$
  - $E(c_V) \to \min$
  - $V(c_V) \to \min$

Apsekojuma dizains



Problēma

- Praksē lietotie izlases dizaini parasti ir sarežģīti
- Tiek lietoti sarežģīti populācijas parametru novērtētāji

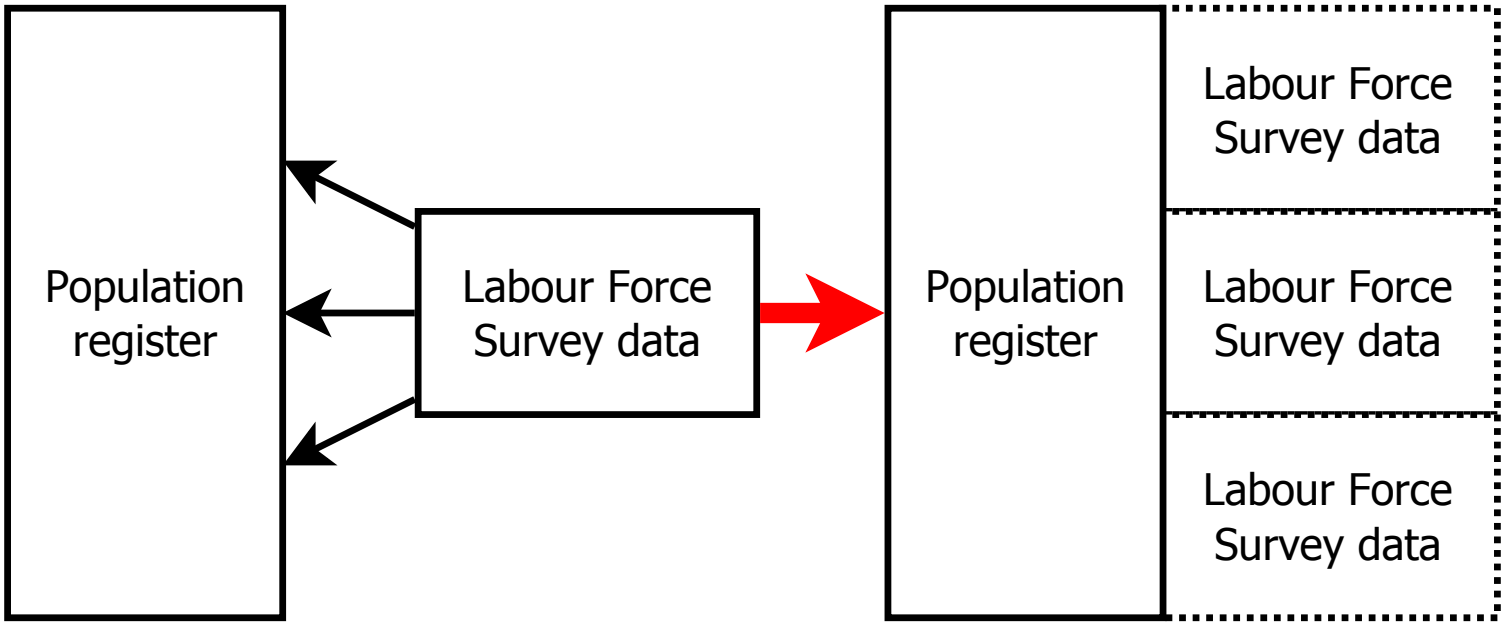
Risinājums

- Iespējams risinājums – apsekojuma dizainu (izlases dizains + novērtēšanas dizains) pētīt ar simulācijas eksperimentu palīdzību
- Ir nepieciešami mākslīgas populācijas dati, lai veiktu simulācijas eksperimentus
- Mākslīgā populācija ir nepieciešama tāda, kas būtu pēc iespējas līdzīga mērķa populācijai. Līdzīga attiecībā pret:
  - informāciju, kas tiek lietota izlases veidošanai un parametru novērtēšanu;
  - ģeogrāfisko izvietojumu;
  - Pētāmajiem rādītājiem.

Mākslīgā populācija

- Mākslīgā populācija tika izveidota, izmantojot:
  - Latvijas Iedzīvotāju reģistra datus;
  - Latvijas Darbaspēka apsekojuma datus.

Mākslīgā populācija

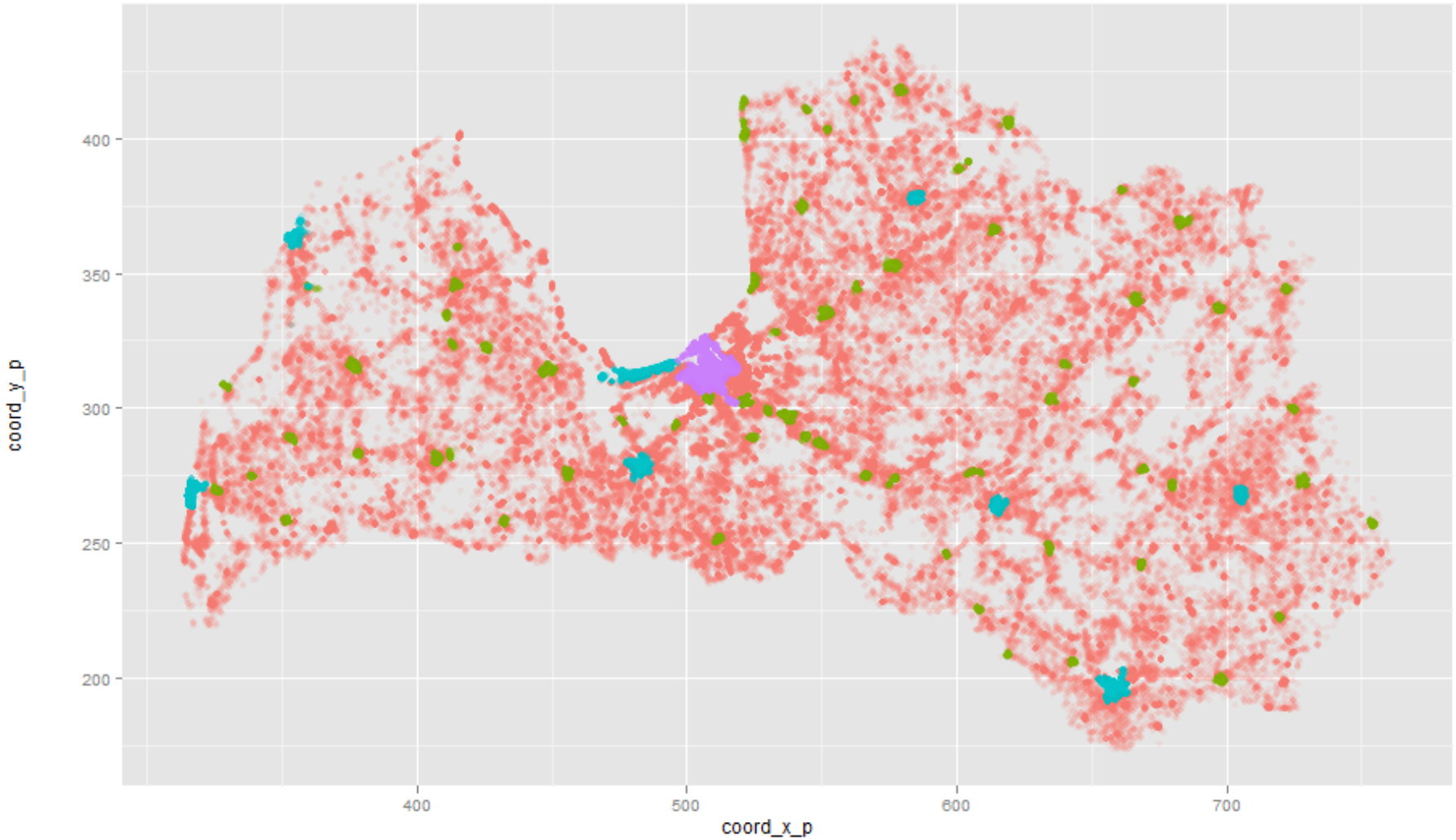


Mākslīgā populācija

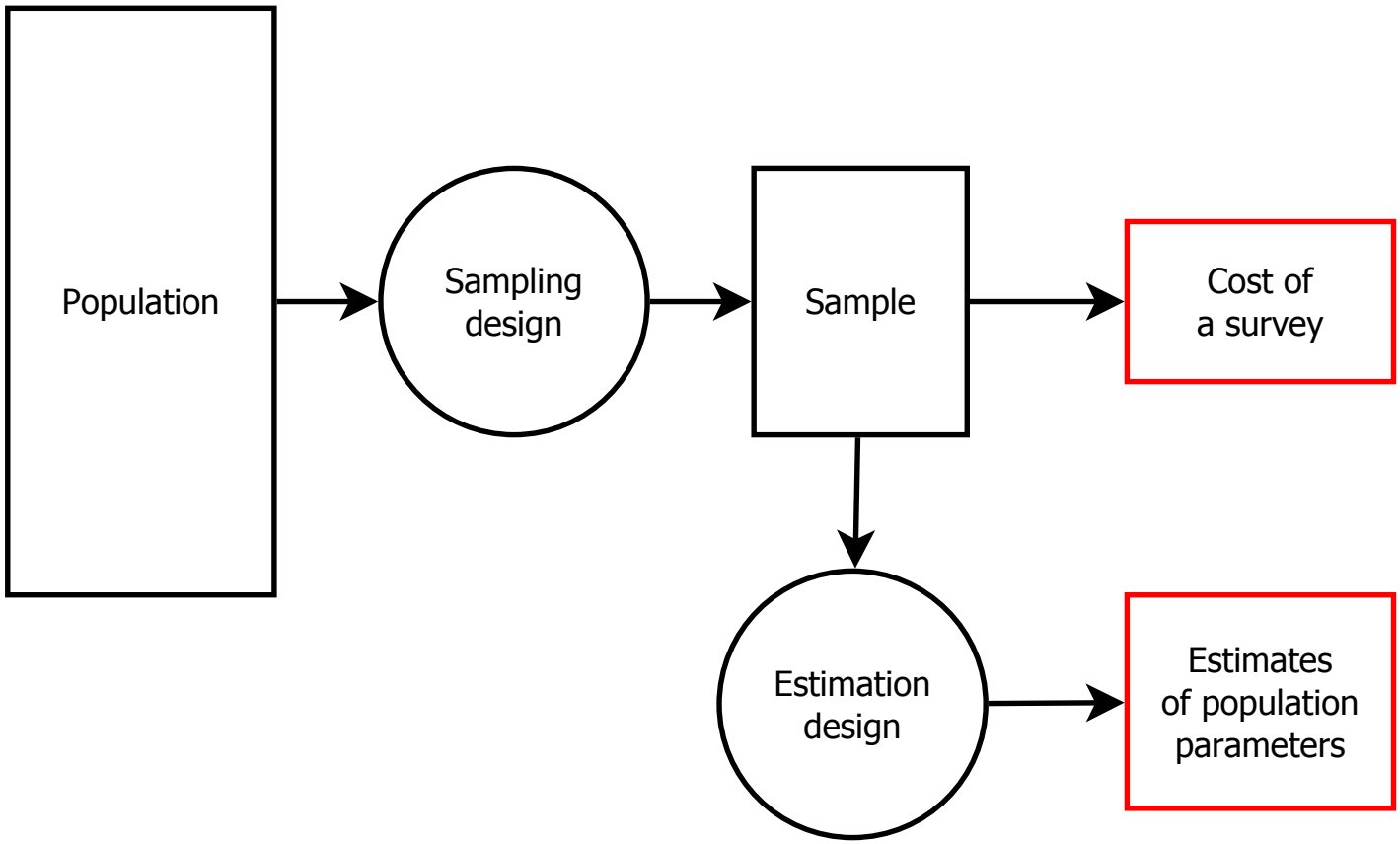
Mākslīgās populācijas raksturojums

- Populācija sastāv no apmēram 1 700 000 personām (765 000 mājsaimniecībām)
- Ģeogrāfiskā novietojuma informācija
- Demogrāfiskā informācija (dzimums, vecums)
- Darbaspēka apsekojuma informācija par ekonomisko aktivitāti un nostrādātajām stundām

Populācijas karte



Apsekojuma dizains

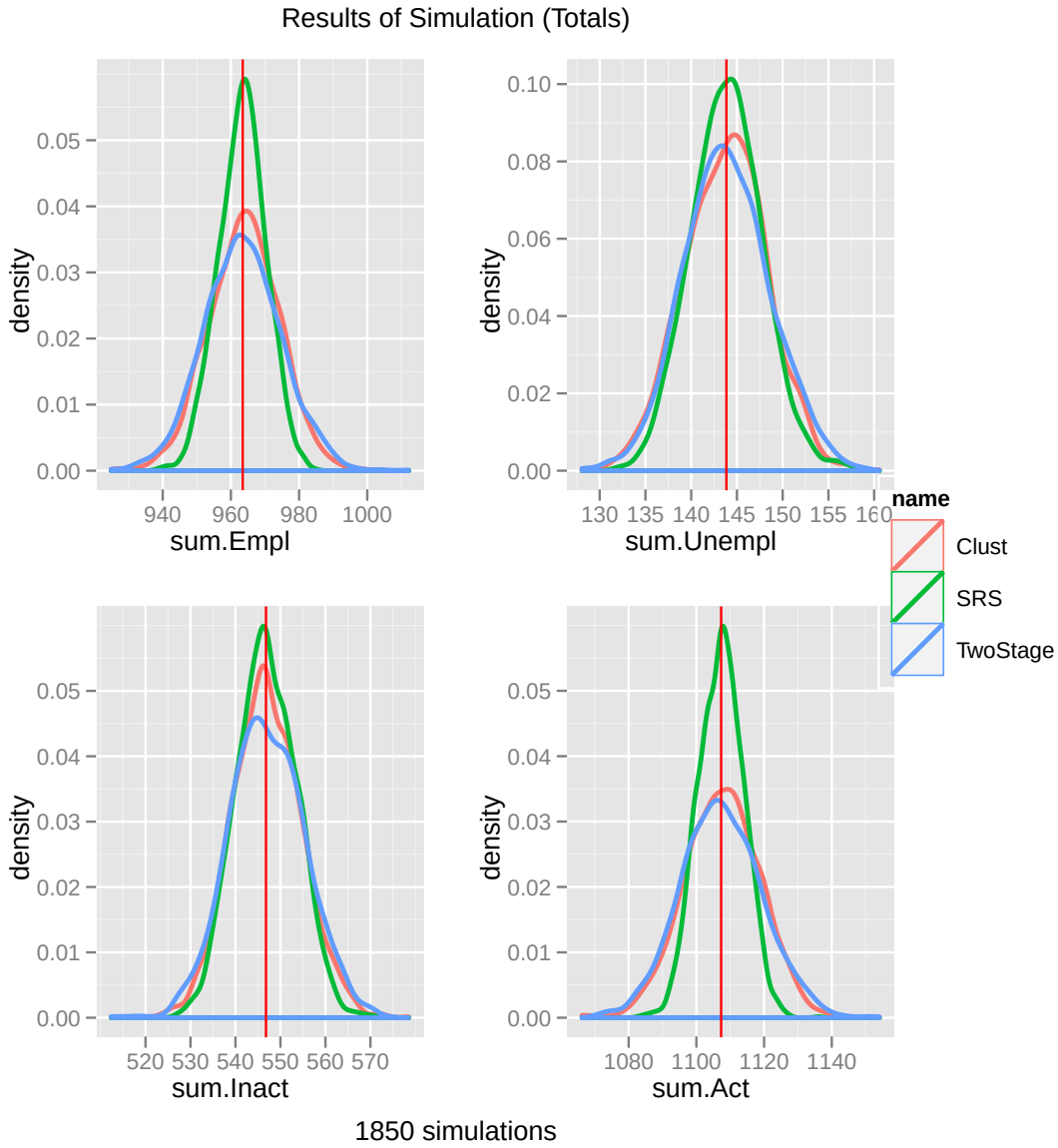


Daži rezultāti

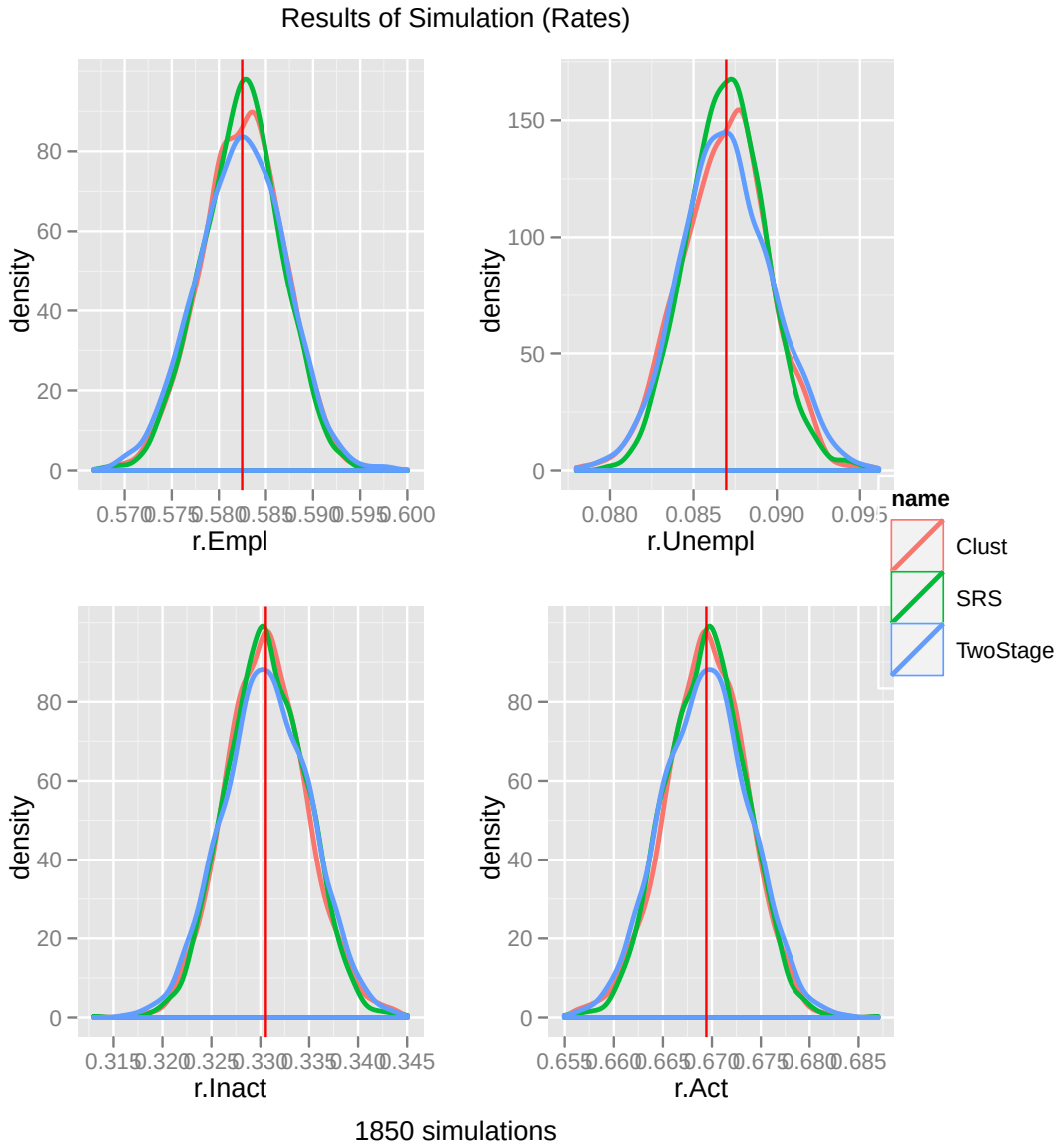
Simulācijas konfigurācija:

- Simulācijā tika izmantoti trīs izlases dizaini:
  - vienkāršā gadījuma izlase no personām ar apjomu 13 413 personas (vidēji vienā mājsaimniecībā ir 2.2 personas);
  - personu klāsterizlase, kur mājsaimniecības veido personu klāsterus, ar apjomu 6 032 klāsteri;
  - divpakāpju izlase, kur PIV ir teritorijas un SIV ir mājsaimniecības (personu klāsteri), ar apjomu 6 032 klāsteri
- HT-novērtētājs
- Simulācijas iterāciju skaits katram izlases dizainam ir 1 850
- Katrā simulācijā tika aprēķināti:
  - populācijas parametru novērtējumi;
  - ceļa garums intervētājiem, kas nepieciešams, lai veiktu apsekojumu

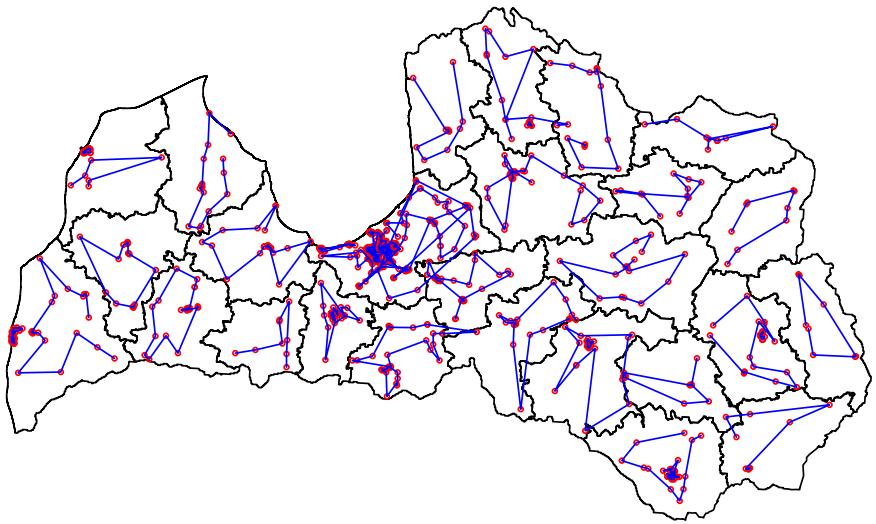
Rezultāti - precizitāte



Rezultāti - precizitāte



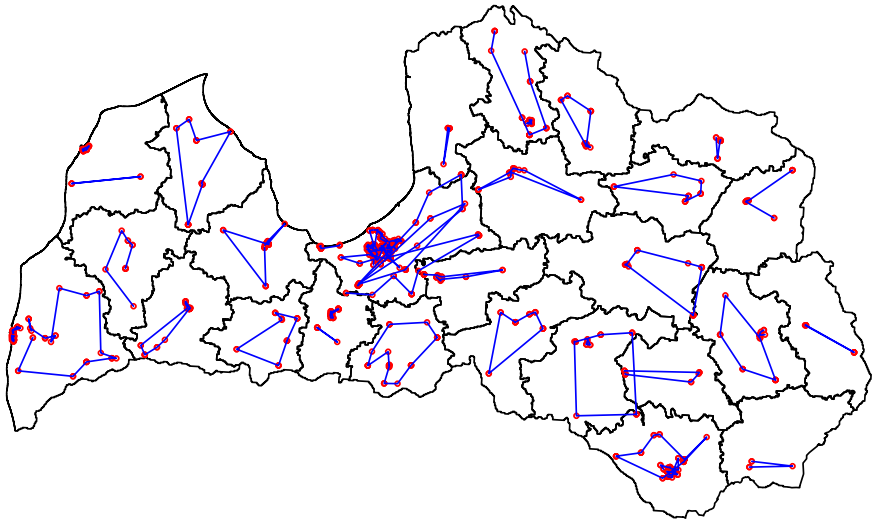
Rezultāti - izmaksas



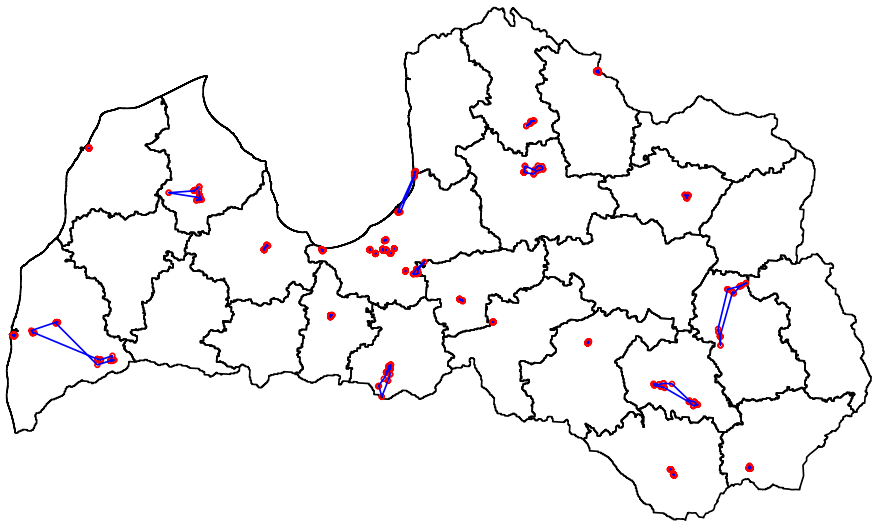
Rezultāti - izmaksas

Cluster Sampling of persons

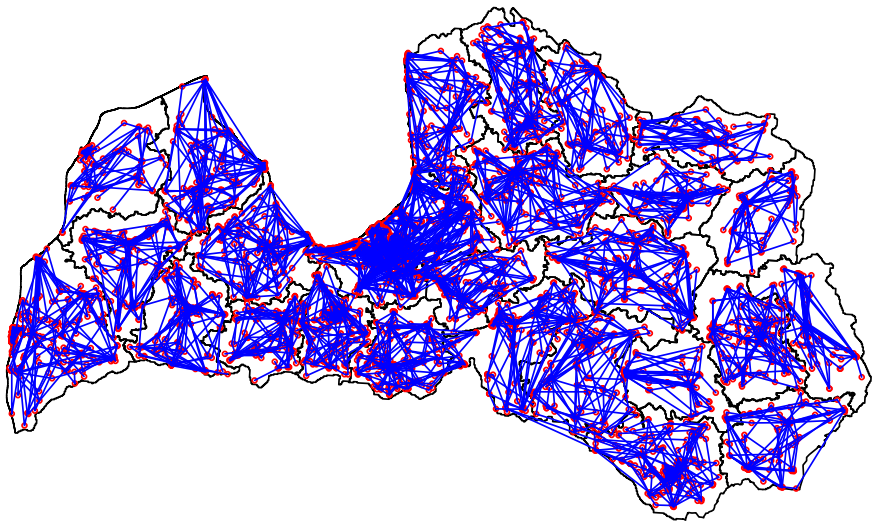
Sample size = 464; Trip = 3242 (km)



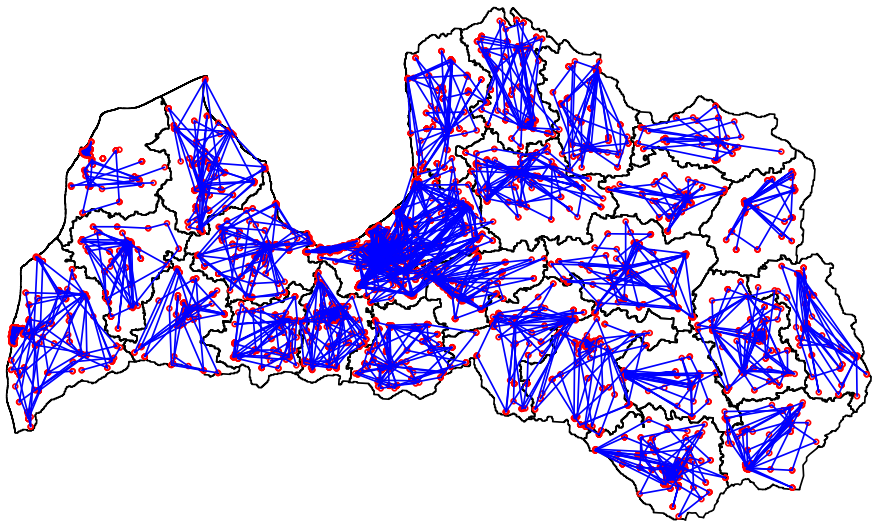
Rezultāti - izmaksas



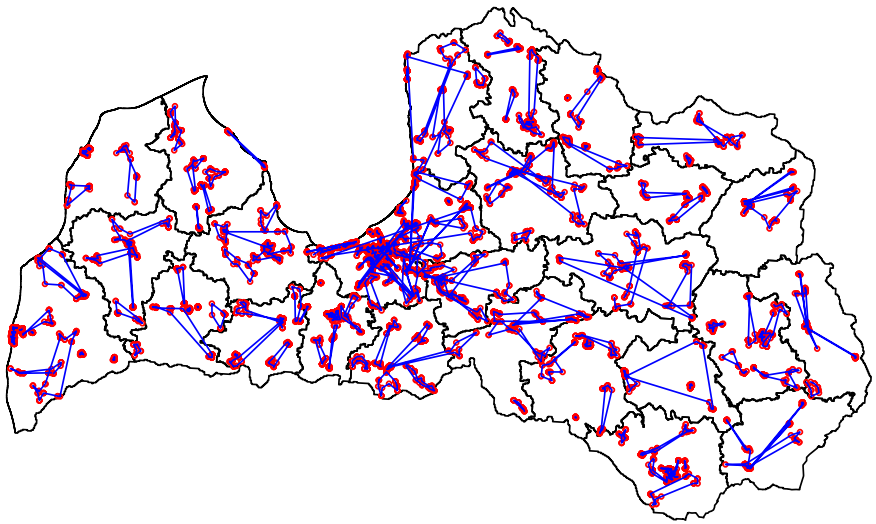
Rezultāti - izmaksas



Rezultāti - izmaksas



Rezultāti - izmaksas



Dizaina efekts

- Tradicionālais dizaina efekts – dizaina efektivitātes mērs pēc precizitātes

$$deff_1 = \frac{Var_{Curr}}{Var_{SRS}}$$

- Varam ieviest alternatīvu dizaina efektu – dizaina efektivitātes mērs pēc izmaksām

$$deff_2 = \frac{Cost_{Curr}}{Cost_{SRS}}$$

Dizaina efekts

- Dizaina efekts – dizaina efektivitātes gan pēc precizitātes, gan pēc izmaksām

$$deff_3 = \frac{Var_{Curr} \cdot Cost_{Curr}}{Var_{SRS} \cdot Cost_{SRS}}$$

Dizaina efekts

design	variable	var	costV	costF	cost	deff1	deff2	deff3
SRS	sum.Act	45	63823	32109	95932	1.000	1.000	1.000
Clust	sum.Act	123	41963	32109	74072	2.737	0.772	2.113
TwoStage	sum.Act	144	8027	32109	40137	3.200	0.418	1.339
SRS	sum.Empl	48	63823	32109	95932	1.000	1.000	1.000
Clust	sum.Empl	108	41963	32109	74072	2.221	0.772	1.715
TwoStage	sum.Empl	128	8027	32109	40137	2.651	0.418	1.109
SRS	sum.Inact	45	63823	32109	95932	1.000	1.000	1.000
Clust	sum.Inact	60	41963	32109	74072	1.337	0.772	1.032
TwoStage	sum.Inact	71	8027	32109	40137	1.573	0.418	0.658
SRS	sum.Unempl	15	63823	32109	95932	1.000	1.000	1.000
Clust	sum.Unempl	20	41963	32109	74072	1.330	0.772	1.027
TwoStage	sum.Unempl	22	8027	32109	40137	1.460	0.418	0.611

Dizaina efekts

design	variable	var	costV	costF	cost	deff1	deff2	deff3
SRS	a.Work.time	0.012294	63823	32109	95932	1.000	1.000	1.000
Clust	a.Work.time	0.015167	41963	32109	74072	1.234	0.772	0.953
TwoStage	a.Work.time	0.016603	8027	32109	40137	1.350	0.418	0.565
SRS	r.Act	0.000016	63823	32109	95932	1.000	1.000	1.000
Clust	r.Act	0.000017	41963	32109	74072	1.059	0.772	0.818
TwoStage	r.Act	0.000020	8027	32109	40137	1.205	0.418	0.504
SRS	r.Empl	0.000018	63823	32109	95932	1.000	1.000	1.000
Clust	r.Empl	0.000019	41963	32109	74072	1.065	0.772	0.823
TwoStage	r.Empl	0.000022	8027	32109	40137	1.229	0.418	0.514
SRS	r.Unempl	0.000006	63823	32109	95932	1.000	1.000	1.000
Clust	r.Unempl	0.000007	41963	32109	74072	1.233	0.772	0.952
TwoStage	r.Unempl	0.000008	8027	32109	40137	1.370	0.418	0.573

Dizaina efekts

- Tradicionālais dizaina efekts – nosacījums: vienāds izlases apjoms

$$deff_1 = \frac{Var_{Curr}\left(n_{Curr} = n\right)}{Var_{SRS}\left(n_{SRS} = n\right)}$$

- Alternatīvs dizaina efektu – nosacījums: vienādas izmaksas

$$deff_4 = \frac{Var_{Curr}\left(c_{Curr} = c\right)}{Var_{SRS}\left(c_{SRS} = c\right)}$$

Secinājumi

- Simulācijas eksperimenti – rīks, lai izvērtētu un analizētu apsekojuma dizaina īpašības:
  - precizitāte
  - izmaksas
- Noderīgs:
  - analizējot apsekojuma efektivitāti
  - plānojot izmaiņas apsekojuma dizainā
  - veicot izmaiņas apsekojuma dizainā

Paldies par uzmanību!