

# Eesti MSMide PSE modelleerimine QGuysi andmete põhjal

Maksym Kasianczuk  
EHPV M&E Koordinaator  
maxim.kasianczuk@gmail.com  
<https://orcid.org/0000-0003-0701-3595>

8.04.2026

Venekeelsete MSMide populatsiooni suuruse hindamiseks (PSE, *population size estimation*) kasutati QGuys veebisaidi profiilide aegrida (Lisa A), milles vaadeldi profiilide arvu muutust ajas eraldi kogu Eesti ja Tallinna kohta.

Sait «QGuys» (2003–2024) määratles end venekeelsete geide ja biseksuaalide suhtlusvõrgustikuna ning soodustas seetõttu kasutajate avalikku suhtlust sõltumata nende geograafilisest paiknemisest. Saidil puudus muu kui venekeelne kasutajaliides, mistõttu said seda kasutada peamiselt vene keelt valdavad isikud. Andmed eri kuupäevadel registreeritud kasutajate arvu kohta pärinevad saidi lehekülgedele koopiatest veebiarhiivis «Wayback Machine».

Andmete modelleerimiseks sobitati kummalegi seeriale klassikaline logistiline kasvumudel, mis võimaldab kirjeldada alguses kiiret kasvu ja sellele järgnevat aeglustumist küllastustaseme lähedal. Mudel esitati kujul [1]:

$$y(t) = \frac{A}{1 + \exp\left(\frac{x_{mid}-t}{s}\right)}, \quad (1)$$

kus  $y(t)$  on profiilide arv ajahetkel  $t$ ,  $A$  on asümptootiline maksimum ehk hinnanguline küllastustase,  $x_{mid}$  on logistilise kõvera keskpunkt ajas ning  $s$  on kasvu kiirust määrav skaalar

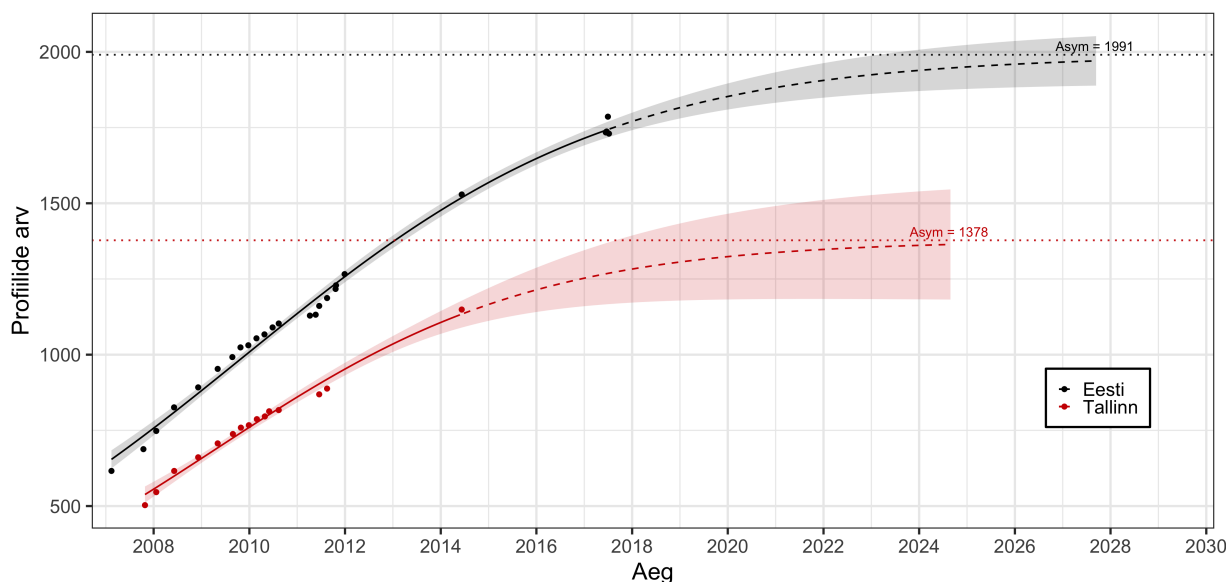
Kõik analüüsid ja arvutused tehti tarkvarakeskkonnas R 4.3.2 [2]. Kasutatud kood on Lisas B.

Parameetrid hinnati mittelineaarse vähimruutude meetodiga. Logistilise mudeli peamine sisuline parameeter on  $A$ , mis näitab profiilide arvu taset, mille poole kõver pika aja jooksul läheneb. Eesti seeria puhul saadi asümptoodiks:  $A_{Eesti} = 1991$ , mille 95% usaldusvahemik (UV) oli [1897–2084]. Tallinna seeria puhul saadi vastavalt  $A_{Tallinn} = 1378$  (joonis 1). Mudelijoont pikendati tulevikku, et visualiseerida kõvera lähenemist asümptoodile; joonisel on vaadeldud perioodi sobitus esitatud täisjoonena ning prognoos katkendjoonena. Varjutatud ala tähistab mudeli hinnangu 95% UV.

Et hinnata venekeelsete MSMide osakaalu Eesti meeste rahvastikus, kasutati Statistikaameti (<https://stat.ee/et>) 2024. aasta andmeid. Venekeelseteks loeti käesolevas analüüsis venelased, valgevenelased ja ukrainlased. Nende osakaal kõigi Eestis elavate meeste hulgas arvutati valemiga:

$$p = \frac{N_{venelased} + N_{valgevenelased} + N_{ukrainlased}}{N_{kõik\ mehed}} \quad (2)$$

Asendades valemisse vastavad väärtused, saame  $p = 0,2592$  ehk 25,92%. Eeldusel, et venekeelsete MSMide osakaal MSMide kogupopulatsioonis vastab nende osakaalule meeste rahvastikus, saab MSMide koguarvu Eestis hinnata valemiga  $PSE = A_{Eesti}/p$  ehk asendades  $A_{Eesti} = 1991$  ja  $p =$



Joonis 1: Venekeelsete MSMide PSE Eestis (logistiline kasvumudel QGuys veebisaidi andmete põhjal)

0,2592, saame 7679 ehk ümardatuna 7700. UV piirid saadi analoogiliselt, jagades venekeelsete MSMide PSE 95% UV piirid sama osakaaluga.

MSMide populatsiooni suuruse hinnanguid on Eestis koostatud alates 2009. aastast [3]. Kõige värskem hinnang (2024) [4] oli 8600 [6000–12100]. Seega ei ole geitutvuste saidi «QGuys» ajalooliste andmete põhjal saadud hinnang ehk 7700 [7300–8000] varasemate hinnangutega vastuolus.

## Viited

- [1] Murray, J. D. *Mathematical Biology I: An Introduction*. 3rd ed. New York: Springer, 2002.
- [2] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2023. <https://www.R-project.org>.
- [3] Kasianczuk, M., Lõhmus, L., Salekešin, M., & Rüütel, K. (2023). *MSM Population Size Estimation in Estonia: Research report*.
- [4] Касянчук, М., Канатова, М., & Милякина, А. (2025). *Квир-партнёрства в странах Балтии*. Таллинн: МТÜ Peemoti Raamatud. 75 с.

## A. Andmed

QGuysi profiilide arv eri kuupäevadel Eestis ja Tallinnas

Kuupäev	Eesti	Tallinn
2007-02-13	616	
2007-10-15	688	
2007-10-27		503
2008-01-21	748	546
2008-06-05	826	
2008-06-06		616
2008-12-06	892	661
2009-05-04	953	707
2009-08-24	992	
2009-08-28		738
2009-10-25	1024	
2009-10-28		759
2009-12-24	1031	
2009-12-28		767
2010-02-24	1054	
2010-02-27		787
2010-04-26	1067	
2010-04-29		796
2010-06-01		813
2010-06-27	1090	
2010-08-13	1103	817
2011-04-08	1129	
2011-05-22	1132	
2011-06-18	1161	869
2011-08-17	1187	888
2011-10-21	1217	
2011-10-24	1229	
2011-12-29	1266	
2014-06-10	1529	1149
2017-06-14	1734	
2017-06-15	1735	
2017-06-20	1737	
2017-06-30	1786	
2017-07-08	1730	

## B. R-kood

```
library(dplyr)
library(tidyr)
library(purrr)
library(ggplot2)
library(minpack.lm)
library(numDeriv)

#-----
# 1. Andmete ettevalmistus
#-----

plot_df <- estoniaqg %>%
```

```

mutate(
  date = as.Date(date),
  seeria = case_when(
    city == "Tallinn" ~ "Tallinn",
    oblast == "В цілому" ~ "Eesti",
    TRUE ~ NA_character_
  )
) %>%
filter(!is.na(seeria)) %>%
select(seeria, date, profiles) %>%
arrange(seeria, date) %>%
group_by(seeria, date) %>%
summarise(profiles = sum(profiles), .groups = "drop")

#-----
# 2. Logistilise mudeli funktsioon
#-----

logistic_fun <- function(t, Asym, xmid, scal) {
  Asym / (1 + exp((xmid - t) / scal))
}

#-----
# 3. Mudeli sobitamise ühe seeria jaoks
#-----

fit_logistic <- function(df_one) {
  origin <- min(df_one$date)

  df_mod <- df_one %>%
    mutate(
      t = as.numeric(date - origin),
      y = profiles
    )

  start_vals <- as.list(
    getInitial(y ~ SSlogis(t, Asym, xmid, scal), data = df_mod)
  )

  model <- nlsLM(
    y ~ SSlogis(t, Asym, xmid, scal),
    data = df_mod,
    start = start_vals,
    control = nls.lm.control(maxiter = 1000)
  )

  list(
    model = model,
    origin = origin
  )
}

#-----
# 4. Prognoos + 95% usaldusvahemik
#-----

predict_logistic_ci <- function(fit_obj, new_dates, level = 0.95) {
  model <- fit_obj$model
  origin <- fit_obj$origin

  t_new <- as.numeric(new_dates - origin)
  cf <- coef(model)
  V <- vcov(model)

  fit <- logistic_fun(
    t = t_new,
    Asym = cf["Asym"],
    xmid = cf["xmid"],
    scal = cf["scal"]
  )

  J <- numDeriv::jacobian(
    func = function(par) {
      logistic_fun(
        t = t_new,
        Asym = par[1],

```

```

      xmid = par[2],
      scal = par[3]
    )
  },
  x = unname(cf)
)

se_fit <- sqrt(pmax(0, diag(J %*% V %*% t(J))))
alpha <- 1 - level
crit <- qt(1 - alpha / 2, df = df.residual(model))

tibble(
  date = new_dates,
  fit = fit,
  lwr = pmax(0, fit - crit * se_fit),
  upr = fit + crit * se_fit
)
}

#-----
# 5. Mudelite sobitamine + tuleviku horisont
#   Prognos kuni ajani, mil kõver jõuab 99% Asym-ist
#-----

forecast_level <- 0.99

models_tbl <- plot_df %>%
  group_by(seeria) %>%
  nest() %>%
  mutate(
    fit_obj = map(data, fit_logistic),
    pred = map2(data, fit_obj, ~{
      cf <- coef(.y$model)

      # aeg (päevades alates origin-ist), mil saavutatakse 99% Asym-ist
      t_target <- unname(cf["xmid"] - cf["scal"] * log(1 / forecast_level - 1))
      date_target <- .y$origin + ceiling(t_target)

      # et prognoosiosa oleks visuaalselt selge, jätame vähemalt 2 aastat pärast viimast vaatlust
      end_date <- max(max(.x$date) + 365 * 2, as.Date(date_target))

      new_dates <- seq(min(.x$date), end_date, by = "month")

      predict_logistic_ci(.y, new_dates) %>%
        mutate(
          seeria = unique(.x$seeria),
          period = if_else(date <= max(.x$date), "Sobitus", "Prognoos")
        )
    })
  )

pred_df <- models_tbl %>%
  select(seeria, pred) %>%
  unnest(pred)

#-----
# 6. Asümptootiliste maksimumide tabel
#-----

asym_tbl <- models_tbl %>%
  mutate(
    coef_vec = map(fit_obj, ~ coef(.x$model))
  ) %>%
  transmute(
    seeria,
    Asym = map_dbl(coef_vec, ~ unname(.x["Asym"]))
  )

print(asym_tbl)

#-----
# 7. Sildid Asym-i jaoks joonisel
#-----

label_df <- pred_df %>%
  group_by(seeria) %>%

```

```

summarise(
  x = max(date) - 250,
  .groups = "drop"
) %>%
left_join(asym_tbl, by = "seeria") %>%
mutate(
  y = Asym,
  silt = paste0("Asym = ", round(Asym, 0))
)

#-----
# 8. Joonis
#-----

ggplot() +
  geom_ribbon(
    data = pred_df,
    aes(x = date, ymin = lwr, ymax = upr, fill = seeria),
    alpha = 0.16,
    colour = NA
  ) +
  geom_line(
    data = pred_df %>% filter(period == "Sobitus"),
    aes(x = date, y = fit, colour = seeria),
    linewidth = 1
  ) +
  geom_line(
    data = pred_df %>% filter(period == "Prognosis"),
    aes(x = date, y = fit, colour = seeria),
    linewidth = 1,
    linetype = "dashed"
  ) +
  geom_point(
    data = plot_df,
    aes(x = date, y = profiles, colour = seeria),
    size = 3
  ) +
  geom_hline(
    data = asym_tbl,
    aes(yintercept = Asym, colour = seeria),
    linewidth = 1.1,
    linetype = "dotted",
    alpha = 0.9
  ) +
  geom_text(
    data = label_df,
    aes(x = x, y = y, label = silt, colour = seeria), show.legend = FALSE,
    hjust = 0.1,
    vjust = -0.4,
    size = 6
  ) +
  scale_colour_manual(
    values = c(
      "Eesti" = "black",
      "Tallinn" = "red3"
    )
  ) +
  scale_fill_manual(
    values = c(
      "Eesti" = "black",
      "Tallinn" = "red3"
    ),
    guide = "none"
  ) +
  scale_x_date(
    date_labels = "%Y",
    date_breaks = "2 years",
    expand = expansion(mult = c(0.02, 0.12))
  ) +
  labs(
    x = "Aeg",
    y = "Profiilide arv",
    colour = NULL,
    title = "Venekeelsete MSMide PSE Eestis\n(logistiline kasvumudel QGuys veebisaidi baasil)",
  )

```

```

caption = str_wrap("Punktid tähistavad vaatlusi; jooned näitavad logistilise mudeli sobitust ja
↳ prognoosi; punktiir tähistab asümptooti ning varjutatud ala 95% UV.\nVenemaalsete MSMide PSE on
↳ 1991 inimest (95% UV 1897-2084). Venemaalsete on 25,92% Eestis elavatest meestest
↳ (Statistikaamet, 2024). Eeldusel, et venemaalsete MSMide osakaal MSMide kogupopulatsioonis
↳ vastab nende osakaalule meeste rahvastikus, võib MSMide koguarvuks Eestis hinnata ligikaudu
↳ 7679 (7317-8042)", width = 120)
) +
theme_bw(base_size = 30) +
theme(
  legend.position = c(0.85, 0.31),
  legend.justification = c(0, 1),
  legend.background = element_rect(fill = "white", colour = "black"),
  plot.caption = element_text(size = 18, hjust = 0)
)

```