

Lexis Diagram

Erick Nasareth

2025-05-02

```
# Bancos de dados

sim <- read.csv2("Data/Obitos_micro_2000-2022_RS.csv")
sinasc <- read.csv2("Data/Nascimentos_micro_2000-2022_RS.csv")

# Função do Gabriel

sim_idade <- function(x) {
  # Função para idade
  if (str_sub(x, 1, 1) < 4) {
    x = 0
  } else if (str_sub(x, 1, 1) == 4) {
    x = 0 + as.numeric(str_sub(x, 2))
  }
  else if (str_sub(x, 1, 1) == 5) {
    x = 100 + as.numeric(str_sub(x, 2))
  }
  else{
    x = NA
  }
  return(x)
}
```

Ajustes nos dados do SIM

```
# Remoção dos NA's nas colunas da data de nascimento e da idade

sim <- sim[!is.na(sim$DTNASC), ]
sim <- sim[!is.na(sim$IDADE), ]

# Ajustando os tipos das variáveis utilizadas

sim <- sim %>%
  mutate(DTNASC = as.Date(DTNASC), DTOBITO = as.Date(DTOBITO), IDADE = as.character(IDADE))

# Utilizando a função do Gabriel para decodificar as idades no banco do SIM

sim$IDADE <- sapply(sim$IDADE, sim_idade)

sim_menor_5 <- sim %>%
  filter(IDADE < 5)
```

```

sim_menor_5 <- sim_menor_5 %>%
  mutate(Cohorte = year(DTNASC), Obito = year(DTOBITO))

sim_menor_5 <- sim_menor_5 %>%
  group_by(Obito, Cohorte, IDADE) %>%
  summarise(obitos = length(Obito), .groups = "keep") %>%
  filter(IDADE <= Obito-Cohorte)

sim_menor_5 <- sim_menor_5 %>%
  arrange(Obito, Cohorte, desc(IDADE))

```

Ajustes nos dados do SINASC

```

sinasc$DTNASC <- ymd(sinasc$DTNASC)
sinasc$DTNASC <- year(sinasc$DTNASC)

# nascidos vivos a cada milhão de nascidos por ano

nascidos_por_ano <- sinasc %>%
  group_by(DTNASC) %>%
  summarise(soma = floor(n()/1000))

```

Questão a. Lexis diagram

- a) Construir o Diagrama de Lexis para os dados de nascidos vivos de 2000 a 2022 da UF escolhida (SINASC) e de óbitos menores de 5 anos (idades simples) para o mesmo período segundo ano de nascimento.

```

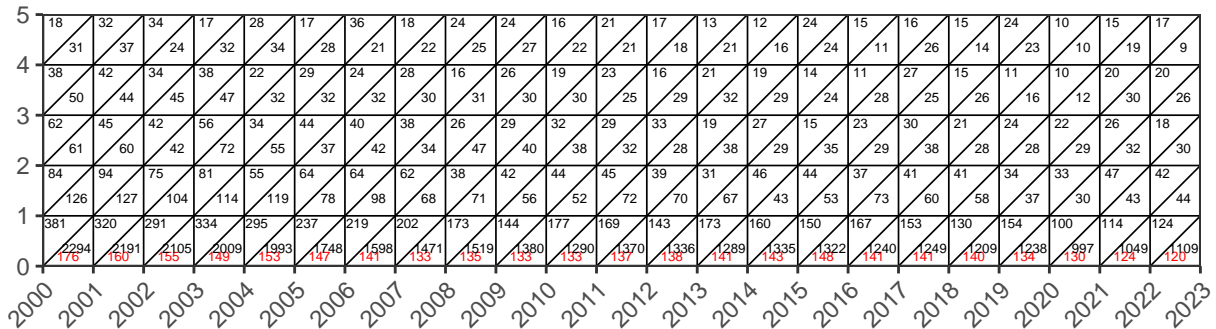
datas <- data.frame(
  anos = as.character(rep(seq(2000, 2022), each = 10)),
  mes = rep(c("04", "09"), 115),
  dia = rep("01", 230)
)

datas <- datas %>%
  mutate(data = as.Date(str_c(anos, "-", mes, "-", dia)))

lexis_grid(year_start = 2000, year_end = 2023, age_start = 0, age_end = 5) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  annotate(
    # Anotações nos gráficos
    "text",
    x = seq(from = as.Date("2000-07-01"), to = as.Date("2022-07-01"), by = "year"),
    y = 0.2,
    label = nascidos_por_ano$soma,
    size = 1.8,
    color = "red"
  ) +
  annotate(
    "text",
    x = datas$data,
    y = rep(seq(4.85, 0, -0.50), 23),
    label = sim_menor_5$obitos,

```

```
size = 1.8
)
```



Questão b. Probabilidade de sobrevivência a idade exata 5

b) Supondo população fechada (inexistência de migração), calcule a probabilidade de um recém-nascido na UF ou território de escolha sobreviver à idade exata 5 para as coortes de 2000 a 2017.

criando o data frame que será utilizado para calcular as probabilidades

```
survive_prob_5yo <- data.frame(
  sim %>%
    filter(IDADE < 5) %>%
    mutate(cohorte = year(DTNASC)) %>%
    filter(cohorte >= 2000 & cohorte <= 2017) %>%
    group_by(cohorte) %>%
    summarise(obitos = n()),
  sinasc %>%
    filter(DTNASC <= 2017) %>%
    group_by(DTNASC) %>%
    summarise(nascidos_vivos = n())
)

survive_prob_5yo <- survive_prob_5yo %>%
  select(cohorte, obitos, nascidos_vivos) %>%
  mutate(probabilidade = 1 - (obitos / nascidos_vivos))
```

```
kable(survive_prob_5yo)
```

cohort	obitos	nascidos_vivos	probabilidade
2000	3034	176719	0.9828315
2001	2898	160590	0.9819540
2002	2805	155261	0.9819336
2003	2671	149165	0.9820937
2004	2548	153015	0.9833480
2005	2288	147199	0.9844564
2006	2076	141331	0.9853111
2007	1921	133401	0.9855998
2008	1905	135143	0.9859038
2009	1803	133652	0.9865097
2010	1708	133243	0.9871813
2011	1764	137710	0.9871905
2012	1730	138941	0.9875487
2013	1693	141350	0.9880226
2014	1714	143315	0.9880403
2015	1733	148359	0.9883189
2016	1598	141411	0.9886996
2017	1593	141568	0.9887475

Questão c. Probabilidade de sobrevivência no primeiro ano de vida

- c) Considerando o mesmo pressuposto, calcule a probabilidade de sobreviver ao primeiro aniversário dos recém-nascidos no período de 2000 a 2020.

```
survive_prob_1yo <- data.frame(
  sim %>%
    filter(IDADE < 1) %>%
    mutate(cohort = year(DTNASC)) %>%
    filter(cohort >= 2000 & cohort <= 2020) %>%
    group_by(cohort) %>%
    summarise(obitos = n()),
  sinasc %>%
    filter(DTNASC <= 2020) %>%
    group_by(DTNASC) %>%
    summarise(nascidos_vivos = n())
)

survive_prob_1yo <- survive_prob_1yo %>%
  select(cohort, obitos, nascidos_vivos) %>%
  mutate(probabilidade = 1 - (obitos / nascidos_vivos))

kable(survive_prob_1yo)
```

cohort	obitos	nascidos_vivos	probabilidade
2000	2614	176719	0.9852082
2001	2482	160590	0.9845445
2002	2439	155261	0.9842910

cohort	obitos	nascidos_vivos	probabilidade
2003	2304	149165	0.9845540
2004	2230	153015	0.9854263
2005	1967	147199	0.9866371
2006	1800	141331	0.9872639
2007	1644	133401	0.9876763
2008	1663	135143	0.9876945
2009	1557	133652	0.9883503
2010	1459	133243	0.9890501
2011	1513	137710	0.9890131
2012	1509	138941	0.9891393
2013	1449	141350	0.9897489
2014	1485	143315	0.9896382
2015	1489	148359	0.9899635
2016	1393	141411	0.9901493
2017	1379	141568	0.9902591
2018	1363	140047	0.9902676
2019	1338	134596	0.9900591
2020	1111	130742	0.9915023