EDA of Endometrial Cancer Incidence in Sweden 1970-2024

Carmen Canedo

```
ec <- read_delim("data/Statistikdatabasen_2026-01-29 10_27_59.csv",
                 # File automatically comes with header. Could be worth deleting manually later
                 skip = 1,
                 delim = ";")
```

```
## Rows: 1210 Columns: 9
## -- Column specification --------------------------------------------------------
## Delimiter: ";"
## chr (1): Region
## dbl (8): År, 017, 1829, 3044, 4564, 6579, 80+, Totalt
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Cleaning

```
ec <- ec %>%
  rename(year = År,
         region = Region,
         age_0_17 = 3,
         age_18_29 = 4,
         age_30_44 = 5,
         age_45_64 = 6,
         age_65_79 = 7,
         age_80_plus = 8,
         age_all = Totalt)
```

```
ec_long <- ec %>%
  pivot_longer(cols = starts_with("age_"),
               names_to = "age",
               values_to = "incidence")
```

Selection Overview

The `ec` table is pulled from Socialstyrelsen Statistikdatabas för cancer with the following selections:

- International incidence definition of morbidity IARC/IACR/ENCR
- Tumor diagnosis (ICD-7): 172 Livmoderkropp (corpus uteri)

  - Note: ICD-10 is C54

- Regardless of tumor type(s)
- All regions, including Riket
- Older age groups (with ages shown separately):
  - 0-17
  - 18-29
  - 30-44
  - 45-64
  - 65-79
  - 80+
- Sex: women
- Years: 1970-2024

Summary Statistics

Descriptive stats for all regions over all years

```
summary_reg_yrs <- ec_long %>%
  summarise(Minimum = min(incidence),
            Mean = mean(incidence),
            Median = median(incidence),
            "Std. Deviation" = sd(incidence, na.rm = TRUE),
            Variance = var(incidence, na.rm = TRUE),
            Max = max(incidence),
            .by = c(year, region)) %>%
  arrange(region)
```

Descriptive stats for all ages regardless of region

```
summary_age_yrs <- ec_long %>%
  summarise(Minimum = min(incidence),
            Mean = mean(incidence),
            Median = median(incidence),
            "Std. Deviation" = sd(incidence, na.rm = TRUE),
            Variance = var(incidence, na.rm = TRUE),
            Max = max(incidence),
            .by = c(year, age)) %>%
  arrange(age)
```

Descriptive stats for

```
# ec %>%
#   ggplot(aes(x = )) +
#   geom_boxplot()
```