

Exercises day 2 – Binomial regression, link functions

- 1) Create a 2x2 table and conduct a binomial regression analysis with a logit link to explore the association between Preterm birth and Passive smoking.

```
# 2x2 frequency table
pregnancies |>
  mutate(
    passive_smoking = factor(
      passive_smoking,
      levels = c(0,1),
      labels = c("No passive smoking",
                 "Passive smoking")
    ),
    preterm_birth = factor(
      preterm_birth,
      levels = c(0,1),
      labels = c("No preterm birth",
                 "Preterm birth")
    )
  ) |>
  count(passive_smoking, preterm_birth) |>
  mutate(
    row_pct = round(100 * n / sum(n), 1),
    display = paste0(n, " (", row_pct, "%)",
      .by = passive_smoking
    ) |>
  select(passive_smoking, preterm_birth, display) |>
  pivot_wider(
    names_from = preterm_birth,
    values_from = display
  )

# Fit regression model
model1 <- glm(
  preterm_birth ~ passive_smoking,
  family = binomial(link = "logit"),
  data = pregnancies
)

# Display results
summary(model1)
```

- a. Is passive smoking a risk factor of preterm birth?

- b. What is the interpretation of the beta coefficient in this analysis?
 - c. How do you calculate the odds ratio from the beta coefficient?
- 2) Repeat step 1 but use a log link in the binomial regression analysis. Interpret the results.
- 3) Repeat step 1 but use an identity link in the binomial regression analysis. Interpret the results.
- 4) Create a 2x4 table and conduct a binomial regression analysis with a logit link to examine the association between preterm birth and pre-pregnancy BMI as a categorical exposure. Interpret the results.

```
# Convert bmi_group to a factor
pregnancies <- pregnancies |>
  mutate(
    bmi_group = factor(
      bmi_group,
      levels = c(1, 2, 3, 4),
      labels = c(
        "Normal",
        "Underweight",
        "Overweight",
        "Obese"
      )
    )
  )

# 2x4 frequency table
pregnancies |>
  count(bmi_group, preterm_birth) |>
  mutate(
    row_pct = round(100 * n / sum(n), 1),
    display = paste0(n, " (", row_pct, "%)",
      .by = bmi_group
    ) |>
  mutate(
    preterm_birth = case_match(
```

```

      preterm_birth,
      0 ~ "No preterm birth",
      1 ~ "Preterm birth"
    )
  ) |>
  select(bmi_group, preterm_birth, display) |>
  pivot_wider(
    names_from = preterm_birth,
    values_from = display
  )

# Fit regression model
model4 <- glm(
  preterm_birth ~ bmi_group,
  family = binomial(link = "logit"),
  data = pregnancies
)

```