

Exercises day 3 – Interactions

For this exercise we have synthetically generated a population of adults aged 50 years and above, with explicitly defined confounding effects and interactions. We then carried out a cross-sectional survey by randomly selecting around 1,000 people from this population. For each person we recorded whether they have ever had a heart attack and 18 lifestyle and health exposures that could potentially be associated with it. The data are synthetic and created for teaching only; the numbers are not real disease rates.

Variable	Meaning
heart_attack	Outcome: 1 = has ever had a heart attack, 0 = no
smoker	1 = current smoker, 0 = non-smoker
high_bp	1 = has high blood pressure (hypertension), 0 = no
family_history	1 = a parent or sibling has had a heart attack, 0 = no
drinks_coffee, vegetarian, left_handed, owns_fridge, night_shift, drinks_tea, uses_internet, married, street_food, plays_sport, owns_motorbike, wears_glasses, has_pet, spicy_food, social_media	15 further lifestyle and household variables, each coded 1 = yes, 0 = no, used in exercise 13

- 1) Import the data, which is comma-separated values (csv) formatted, and check that it has been correctly imported.

```
# Load packages
library(tidyverse)

# Set working directory
setwd("__directory path__")

# Load the .csv file
SurveyData <- read_csv("SurveyData.csv")

# Check the import
glimpse(SurveyData)
summary(SurveyData)
```

You should see 19 yes/no (0/1) variables and 1,065 observations. The outcome is heart_attack (1 = yes, 0 = no).

From past studies we know that smoking and high blood pressure may be associated with heart attack. In the data these are the variables smoker, high_bp and heart_attack. We will examine these associations on an additive scale (risk differences) and on a logarithmic scale (relative risks).

- 2) Perform separate univariate binomial regression analyses with a log link for smoker and for high_bp.

```
model_smoker_log <- glm(heart_attack ~ smoker, family = binomial(link = "log"), data = SurveyData)
```

```
model_bp_log <- glm(heart_attack ~ high_bp, family = binomial(link = "log"), data = SurveyData)
```

```
summary(model_smoker_log)
summary(model_bp_log)
```

```
round(exp(cbind(
  RR = coef(model_smoker_log),
  confint(model_smoker_log))),3)
```

```
round(exp(cbind(
  RR = coef(model_bp_log),
  confint(model_bp_log))),3)
```

- a. Are smoking and high blood pressure associated with heart attack? What do the intercept and the coefficient mean with a log link?

3) Smoking is our main exposure. Add high blood pressure to the model to check whether it confounds the association between smoking and heart attack.

```
model_both_log <- glm(
  heart_attack ~ smoker + high_bp,
  family = binomial(link = "log"),
  data = SurveyData
)
summary(model_both_log)
```

- a. Does it look like high blood pressure confounds the association of smoking with heart attack?

4) Estimate the interaction coefficient using the observed risks in the four groups.

$$\text{Interaction}_{\text{multiplicative}} = \frac{R_{\text{Smoker, High BP}} / R_{\text{Non-smoker, High BP}}}{R_{\text{Smoker, Normal BP}} / R_{\text{Non-smoker, Normal BP}}}$$

Interaction coefficient = (Relative Risk of smoking in high BP) / (Relative Risk of smoking in normal BP)

Run below R commands

```
# Risk in the four exposure groups
table_risk <- SurveyData |>
mutate(
  smoker = ifelse(smoker == 1, "Smoker", "Non-smoker"),
  high_bp = ifelse(high_bp == 1, "High BP", "Normal BP")
) |>
```

```

summarise(
  n = n(),
  risk = mean(heart_attack),
  .by = c(smoker, high_bp)
)

R_Smoker_HighBP <- table_risk |>
  filter(smoker == "Smoker", high_bp == "High BP") |>
  pull(risk)

R_NonSmoker_HighBP <- table_risk |>
  filter(smoker == "Non-smoker", high_bp == "High BP") |>
  pull(risk)

R_Smoker_NormalBP <- table_risk |>
  filter(smoker == "Smoker", high_bp == "Normal BP") |>
  pull(risk)

R_NonSmoker_NormalBP <- table_risk |>
  filter(smoker == "Non-smoker", high_bp == "Normal BP") |>
  pull(risk)

beta_interaction <- (R_Smoker_HighBP/R_NonSmoker_HighBP)/
  (R_Smoker_NormalBP / R_NonSmoker_NormalBP)

table_risk
beta_interaction

```

- 5) Estimate the interaction coefficient using the generalized linear model with an interaction term on a multiplicative scale (log link) and compare with the additive scale (identity link)

```

model_int <- glm(
  heart_attack ~ smoker * high_bp,
  family = binomial(link = "identity"),
  data = SurveyData
)
summary(model_int)

model_log <- glm(
  heart_attack ~ smoker * high_bp,
  family = binomial(link = "log"),
  data = SurveyData
)
summary(model_log)

round(exp(cbind(
  RR = coef(model_log),
  confint(model_log))),3)

```

- a. What is the difference between the interaction coefficients obtained from the two approaches?
- b. Is there an interaction between smoking and blood pressure?
- c. Why is there an interaction on the additive scale but not on the multiplicative scale?

- 6) Repeat the interaction model with a logit link (odds ratios). Then put the three scales side by side and compare how the interaction effect and significance changes with scale

```
model_logit <- glm(
  heart_attack ~ smoker * high_bp,
  family = binomial(link = "logit"),
  data = SurveyData
)
summary(model_logit)

interaction_table <- data.frame(
  Scale = c(
    "Additive (identity), RD",
    "Multiplicative (log), RR",
    "Odds (logit), OR"
  ),
  `Interaction-term` = c(
    coef(model_int)["smoker:high_bp"],
    exp(coef(model_log)["smoker:high_bp"]),
    exp(coef(model_logit)["smoker:high_bp"])
  ),
  `P-value` = c(
    coef(summary(model_int))["smoker:high_bp", "Pr(>|z|)"],
    coef(summary(model_log))["smoker:high_bp", "Pr(>|z|)"],
    coef(summary(model_logit))["smoker:high_bp", "Pr(>|z|)"]
  ),
  check.names = FALSE
)

interaction_table[, 2:3] <- round(interaction_table[, 2:3], 5)
interaction_table
```