Exercise 1

A study looking at breast cancer in women compared cases with non-cases, and found that $75/100$ cases did not use calcium supplements compared with $25/100$ of the non-cases.

(a) Develop a 2 x 2 table to display the data.

	Case ("Diseased")	Control ("Healthy")
Exposed	25	75
Not Exposed	75	25

odds ratios
=! odds, ratio of odds

OR = (25/75)/(75/25) = 1/9

(b) Calculate the odds ratio (OR) based on the 2×2 table.

(c) Interpret the odds ratio.

(d) Define the following data frame in R and fit the binary logistic regression model by using the below `glm()` function.

```
disease <- factor(c(rep(1, 75), rep(0, 25), rep(1, 25), rep(0, 75)),
                  levels = c("1", "0"),
                  labels = c("yes", "no"))
supp <- factor(c(rep("no.calc", 100), rep("calc", 100)),
                   levels = c("no.calc", "calc"),
                   labels = c("no.calc", "calc"))
RCT <- data.frame(disease = disease,
                  supp = supp)
table(RCT$supp, RCT$disease)
bin.log.mod <- glm(disease ~ supp, family = "binomial", data = RCT)
```

exp(2.197) = 9, because the logistic regression gives us log odds.

(e) How can you extract the (same) odds ratio as calculated in b) from the `glm()` model?

Exercise 2

The `CHFLS` data set from package **HSAUR2** is a subset of the Chinese Health and Family Life Survey (cf. Exercise 3, Worksheet 1).

We want to study the impact of age and income on happiness in a proportional odds logistic model. For this, the following `polr()` model is fitted:

```
library("MASS")
data("CHFLS", package = "HSAUR2")
polr.mod <- polr(R_happy ~ R_age + R_income, data = CHFLS)
```

(a) What are the characteristics of the response variable `R_happy`?

Apply the following commands to the previously defined R object `polr.mod`.

```
summary(polr.mod)
c(polr.mod$zeta, coef(polr.mod))
exp(c(polr.mod$zeta, coef(polr.mod)))
```

(b) Do the coefficients from a porportional odds logistic regression model have a multiplicative or an additive effect?

(c) Interpret $\exp(\beta_{\text{age}})$ and $\exp(\beta_{\text{income}})$.

(d) How do you interpret the `polr.mod$zeta` coefficients?

Exercise 3

The model studying the impact of age and income on happiness from the previous exercise can also be used to predict outcome probabilities.

```
data("CHFLS", package = "HSAUR2")
library("MASS")
polr.mod <- polr(R_happy ~ R_age + R_income, data = CHFLS)
predict(polr.mod, type = "prob",
        newdata = data.frame(R_age = mean(CHFLS$R_age),
                             R_income = mean(CHFLS$R_income)))
```

(a) Based on the `polr.mod`, predict the probabilities of the self-reported happiness groups (`R_happy`) by changing the values of `R_age` while keeping the `R_income` variable constant.

```
age <- c(min(CHFLS$R_age), mean(CHFLS$R_age), max(CHFLS$R_age))
nd.age <- expand.grid(R_age = age, R_income = mean(CHFLS$R_income))
```

(b) Based on the `polr.mod`, predict the probabilities of the self-reported happiness groups (`R_happy`) by changing the values of `R_income` while keeping the `R_age` variable constant.

```
income <- c(min(CHFLS$R_income), mean(CHFLS$R_income), max(CHFLS$R_income))
nd.income <- expand.grid(R_age = mean(CHFLS$R_age), R_income = income)
```

(c) Are the predicted probabilities in alignment with the regression model coefficients? Why?

Exercise 4

Show that

(a) In the model where

$$F_Y(y) = \mathbb{P}(Y \leq y) = \text{expit}(h(y)),$$

it holds that

$$h(y) = \log\left(\frac{F(y)}{1 - F(y)}\right).$$

(b) In the model where

$$F_Y(y) = \mathbb{P}(Y \leq y) = 1 - \exp(-\exp(h(y))),$$

it holds that

$$h(y) = \log\left(\Lambda(y)\right),$$

where $\Lambda(y)$ is the *cumulative hazard function*.

luisa.barbanti@uzh.ch
lucasheinrich.kook@uzh.ch