

Duration of Unemployment - Different Codings of Covariables

August 21, 2024

The unemployment data represent a contingency table with rows referring to gender and columns to duration of unemployment.

```
> unemployment <- matrix(c(403, 238, 167, 175), nrow=2, ncol=2)
> rownames(unemployment) <- c("male", "female")
> colnames(unemployment) <- c("<6 month", ">6 month")
> unemployment
> rowSums(unemployment)
```

Calculation of odds and log-odds.

```
> ( odds_m <- 403/167 )
> ( odds_w <- 238/175 )
> ( log_odds_m <- log(403/167) )
> ( log_odds_w <- log(238/175) )
```

For the fitting of a logit-model an alternative dataset is generated. First (0-1)-coding is considered

```
> gender <- c(rep(1, 403+167), rep(0, 238+175))
> unemp <- c(rep(1, 403), rep(0, 167), rep(1, 238), rep(0, 175))
```

For control, one can compute the crosstabulation of the generated data.

```
> table(gender, unemp)
```

Fit of a logit model.

```
> bin <- glm(unemp ~ gender, family=binomial)
> summary(bin)
> bin$coef
> exp(bin$coef)
```

Now a dataset in effect-coding is created.

```
> gender_effect <- c(rep(1, 403+167), rep(-1, 238+175))
```

For control, one can compute the crosstabulation of the generated data.

```
> table(gender_effect, unemp)
```

Fit a logit model.

```

> bin_effect <- glm(unemp ~ gender_effect, family=binomial)
> summary(bin_effect)
> bin_effect$coef
> exp(bin_effect$coef)

```

Now we consider education level as explanatory variable.

```

> unemp_level <- matrix(c(202, 307, 87, 45,
+                        96, 162, 66, 18), nrow=4, ncol=2)
> colnames(unemp_level) <- c("Short term", "Long term")
> unemp_level
> rowSums(unemp_level)

```

For the fitting of a logit-model a new dataset is generated. First (0-1)-coding is considered.

```

> level <- factor(c(rep(1, 202+96), rep(2, 307+162), rep(3, 87+66), rep(4, 45+18)))
> unemp_1 <- c(rep(1, 202), rep(0, 96), rep(1, 307), rep(0, 162),
+             rep(1, 87), rep(0, 66), rep(1, 45), rep(0, 18))

```

For control, one can compute the crosstabulation of the generated data.

```

> table(level, unemp_1)

```

Fit a logit model on the data. Define the variable level as a factor with the reference category 4.

```

> level <- relevel(level, ref=4)
> bin_1 <- glm(unemp_1 ~ level, family=binomial)
> summary(bin_1)

```

Now additionally quasi-variances can be computed. Therefore the function "qvcalc" from the "qvcalc"-library is used.

```

> library(qvcalc)
> qv<-qvcalc(bin_1, "level")
> summary(qv)
> plot(qv)

```