Eli Lauwers - Causality Project

2022-07-24

Background

Copy background

Preparation

Read the dataset and modify it as follows:

(code available in the .Rmd file)

In our analysis of this observational study we will aim to evaluate the marginal effect of quitting smoking (the information is held in the variable `qsmk`) on the weight gain between years 1971 and 1982 (the information is held in the variable `wt82_71_bin`, which is 1, in case of weight gain and 0 otherwise). The tutorial [1] may be a helpful reference for your analysis. Throughout, you may assume that the dataset contains a collection of covariates that is sufficient to control for confounding of the effect of quitting smoking on the risk of weight gain.

Logistic Regression outcome model

Estimate the marginal effect of quitting smoking on the risk of weight gain as a risk difference and a relative risk using a logistic regression outcome model. There is no need to be very exhaustive in building a model. You may for instance consider using a standard model choice, and make use of automated procedures (e.g. step in R) to select variables, if needed. More thorough model building will be considered later using causal machine learning methods.

```
source("step_backwards.R")
model = step_backwards(data = nhefs.nmv, outcome = "wt82_71_bin", treatment = "qsmk",
    verbose = 1)
```

```
Predictors in backwards procedure:
   Kept: qsmk, dbp, sex, age, wt71, smokeintensity, active
   Removed: race, smokeyrs, education, tax71_82, sbp, exercise
```

```
p0 = mean(predict(model, na.omit(mutate(nhefs.nmv, qsmk = 0)), type = "response"))
p1 = mean(predict(model, na.omit(mutate(nhefs.nmv, qsmk = 1)), type = "response"))
cat(c("Risk Difference\t", p1 - p0, "\n"))
```

```
Risk Difference  0.127409494250903
```

```
cat(c("Relative risk\t", p1/p0, "\n"))
```

```
Relative risk    1.19893111224333
```

```
model.std = stdReg::stdGlm(model, data = nhefs.nmv, X = "qsmk", x = c(0, 1))
summary(model.std, contrast = "difference", reference = 0)
```

```
Formula: wt82_71_bin ~ qsmk + dbp + sex + age + wt71 + smokeintensity +
    active
<environment: 0x00000000150fab70>
Family: binomial
Link function: logit
Exposure:  qsmk
Reference level:  qsmk = 0
Contrast:  difference

  Estimate Std. Error lower 0.95 upper 0.95
0    0.000     0.0000     0.0000      0.000
1    0.128     0.0238     0.0809      0.174
```

```
summary(model.std, contrast = "ratio", reference = 0)
```

```
Formula: wt82_71_bin ~ qsmk + dbp + sex + age + wt71 + smokeintensity +
    active
<environment: 0x00000000150fab70>
Family: binomial
```

```
Link function: logit
Exposure:  qsmk
Reference level:  qsmk = 0
Contrast:  ratio

  Estimate Std. Error lower 0.95 upper 0.95
0      1.0     0.0000       1.00       1.00
1      1.2     0.0398       1.12       1.28
```