Combining Coefficients and Variances

Ashley I Naimi

Spring 2024

Contents

1 Combining Parameter Estimates from a Regression Model

Suppose we are fitting a log-linear or linear model to estimate the risk difference, risk ratio, or some other contrast on the additive or multiplicative scale, and we have to include an interaction between between the exposure of interest and a covariate in the model. For example:

$$E(Y \mid X, C) = \beta_0 + \beta_1 X + \beta_2 C + \beta_3 XC$$

To be specific, let's consider the NHEFS data again:

```
packages <- c("broom","here","tidyverse","skimr","rlang","sandwich","boot", "kableExtra")

pacman::p_load(packages)

#' Define where the data are
file_loc <- url("https://bit.ly/47ECRcs")

#' This begins the process of cleaning and formatting the data
nhefs <- read_csv(file_loc) %>%
  select(qsmk,wt82_71,exercise,sex,age,
         race,income, marital,school) %>%
  mutate(income=as.numeric(income>15),
         marital=as.numeric(marital>2)) %>%
  na.omit(.)

factor_names <- c("exercise","income","marital","sex","race")
nhefs[,factor_names] <- lapply(nhefs[,factor_names] , factor)

#' Define outcome
nhefs <- nhefs %>% mutate(id = row_number(),
                          wt_delta = as.numeric(wt82_71>median(wt82_71)),
                          .before = qsmk)

#' Quick summary of data
nhefs %>% print(n=5)
```

```
## # A tibble: 1,507 x 11
##      id wt_delta  qsmk wt82_71 exercise sex     age race  income marital school
##   <int>    <dbl> <dbl>   <dbl> <fct>    <fct> <dbl> <fct> <fct>  <fct>    <dbl>
## 1     1        0     0  -10.1  2        0        42 1     1      0            7
## 2     2        1     0    2.60 0        0        36 0     1      0            9
## 3     3        1     0    9.41 2        1        56 1     0      1           11
## 4     4        1     0    4.99 2        0        68 1     0      1            5
## 5     5        1     0    4.99 1        0        40 0     1      0           11
## # i 1,502 more rows
```

Let's fit a linear risk model to estimate the effect of quitting smoking on greater than median weight change,

```
library(lmtest)
library(sandwich)

mod <- lm(wt_delta ~ qsmk + sex + qsmk*sex + race + income,
          data = nhefs)

summary(mod)
```

```
##
## Call:
## lm(formula = wt_delta ~ qsmk + sex + qsmk * sex + race + income,
##     data = nhefs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6274 -0.4791 -0.3930  0.5098  0.6070
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.42512    0.03698  11.496   <2e-16 ***
## qsmk         0.09711    0.04112   2.361   0.0183 *
```

```
## sex1         0.01108    0.02975   0.372   0.7097
## race1       -0.03216    0.03928  -0.819   0.4130
## income1      0.05399    0.03389   1.593   0.1114
## qsmk:sex1    0.04006    0.05922   0.676   0.4988
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4975 on 1501 degrees of freedom
## Multiple R-squared:  0.0138, Adjusted R-squared:  0.01051
## F-statistic: 4.199 on 5 and 1501 DF,  p-value: 0.0008565
```

```r
# we'll need robust SEs:
mod_res <- coeftest(mod, vcov = vcovHC(mod, type = "HC3"))

mod_res
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.425120   0.037303 11.3965  < 2e-16 ***
## qsmk         0.097111   0.041244  2.3546  0.01867 *
## sex1         0.011076   0.029925  0.3701  0.71135
## race1       -0.032164   0.039278 -0.8189  0.41299
## income1      0.053987   0.034137  1.5815  0.11398
## qsmk:sex1    0.040061   0.058844  0.6808  0.49610
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on this model, we might interpret the stratum specific `qsmk` effects separately. The risk difference among individuals classified as `male` (sex = 0) is simply the coefficient for `qsmk` in the model. The risk difference among individuals classified as `female` (sex = 1) is the coefficient for `qsmk` in the model plus the coefficient for `qsmk*sex`. There are a few ways we can compute this in R.

1.1 Method 1

```
rd_sex0_m1 <- mod_res[2,1]

rd_sex1_m1 <- mod_res[2,1] + mod_res[6,1]
```

1.2 Method 2

```
rd_sex0_m2 <- mod_res["qsmk",1]

rd_sex1_m2 <- mod_res["qsmk",1] + mod_res["qsmk:sex1",1]
```

1.3 Method 3

```
contrast1 <- c(0, 1, 0, 0, 0, 0)

contrast2 <- c(0, 1, 0, 0, 0, 1)

rd_sex0_m3 <- coef(mod_res) %*% contrast1

rd_sex1_m3 <- coef(mod_res) %*% contrast2
```

All these approaches return the same risk differences:

Table 1: Comparison of risk differences from all three methods.

Method 1	0.0971114653013959	0.137172888772212
Method 2	0.0971114653013959	0.137172888772212
Method 3	0.0971114653013959	0.137172888772212

2 Computing Standard Errors for Combined Parameters

The next question is how to obtain the correct standard errors for both of these estimates. We've already computed the robust variance estimates for the coefficients. We can thus obtain the standard error for the `qsmk` coefficient directly from the model. The first two methods are easy:

```
var_rd_sex0_m1 <- mod_res[2,2]

var_rd_sex0_m2 <- mod_res["qsmk",2]
```

The third method will require a little explaining, but is very useful to know in more complicated settings. Consider the vector of coefficients from the model above, and let's call it β. In our case, β is a vector with 6 elements.

This vector can be used to define a 6×6 variance-covariance matrix. The diagonal entries of this matrix are the variances of the coefficients from the model. The square root of these are the standard errors:

```
round(vcovHC(mod, type = "HC3"),4)
```

```
##             (Intercept)    qsmk   sex1   race1 income1 qsmk:sex1
## (Intercept)      0.0014 -0.0005 -5e-04 -0.0005 -0.0010    0.0005
## qsmk            -0.0005  0.0017  5e-04  0.0001  0.0000   -0.0017
## sex1            -0.0005  0.0005  9e-04  0.0000  0.0001   -0.0009
## race1           -0.0005  0.0001  0e+00  0.0015  0.0003    0.0000
## income1         -0.0010  0.0000  1e-04  0.0003  0.0012    0.0000
## qsmk:sex1        0.0005 -0.0017 -9e-04  0.0000  0.0000    0.0035
```

```
sqrt(vcovHC(mod, type = "HC3")[2,2]) == mod_res["qsmk",2]
```

```
## [1] TRUE
```

Thus, to obtain the standard error for the `qsmk` parameter with the third method, we can extract the variance for the coefficient from this variance covariance matrix:

```
var_rd_sex0_m3 <- sqrt(contrast1%*%vcovHC(mod, type = "HC3")%*%contrast1)
```

To obtain the standard error for the combined parameter, we have to note that the following variance equation:

$$Var(A + B) = var(A) + var(B) + 2cov(A, B)$$

Because of this equation, we can't simply use the output of the `coeftest` function since we do not have the covariances available in this output. However, the covariances are available from the `vcov` function:

```
round(vcovHC(mod, type = "HC3"),4)
```

```
##             (Intercept)    qsmk    sex1   race1 income1 qsmk:sex1
## (Intercept)      0.0014 -0.0005 -5e-04 -0.0005 -0.0010    0.0005
## qsmk            -0.0005  0.0017  5e-04  0.0001  0.0000   -0.0017
## sex1            -0.0005  0.0005  9e-04  0.0000  0.0001   -0.0009
## race1           -0.0005  0.0001  0e+00  0.0015  0.0003    0.0000
## income1         -0.0010  0.0000  1e-04  0.0003  0.0012    0.0000
## qsmk:sex1        0.0005 -0.0017 -9e-04  0.0000  0.0000    0.0035
```

We can combine the requisite pieces according to the equation above in a couple of ways. For example:

```
var_rd_sex1_m2 <- sqrt(vcovHC(mod, type = "HC3")[2,2] +
                         vcovHC(mod, type = "HC3")[6,6] +
                         2*vcovHC(mod, type = "HC3")[2,6])
```

Alternatively, we can use matrix multiplication to simplify the math:

```
var_rd_sex1_m3 <- sqrt(contrast2%*%vcovHC(mod, type = "HC3")%*%contrast2)
```

Here are the standard errors from each of the methods:

Table 2: Comparison of standard errors from all three methods.

Method 1	0.0412441687682876	-
Method 2	0.0412441687682876	0.0420913863125778
Method 3	0.0412441687682876	0.0420913863125778

With these, we can construct the stratum specific risk difference and 95% confidence intervals:

```
knitr::kable(caption = "Risk differences and 95% CIs using standard errors obtained from method 3.",
  rbind(c("Stratum Sex = 0", round(rd_sex0_m3, 3),
                             round(rd_sex0_m3 - 1.96*var_rd_sex0_m3, 3),
                             round(rd_sex0_m3 + 1.96*var_rd_sex0_m3, 3)),
        c("Stratum Sex = 1", round(rd_sex1_m3, 3),
                             round(rd_sex1_m3 - 1.96*var_rd_sex1_m3, 3),
                             round(rd_sex1_m3 + 1.96*var_rd_sex1_m3, 3))),
  "simple"
)
```

Table 3: Risk differences and 95% CIs using standard errors obtained from method 3.

Stratum Sex = 0	0.097	0.016	0.178
Stratum Sex = 1	0.137	0.055	0.22