# HMRK 14 - Beimnet Taye

2023-05-06

## P1

### 1

```
study_data = medicaldata::indo_rct
fit <- glm(outcome~rx, study_data, family = binomial)
avg_comparisons(model = fit,variables = "rx",vcov = "HC3",
                comparison =  \(hi, lo) (mean(hi)/(1 - mean(hi)))/(mean(lo)/(1 - mean(lo))) ,
                hypothesis = 1)
```

```
##
##  Term                  Contrast Estimate Std. Error     z Pr(>|z|) 2.5 % 97.5 %
##    rx 1_indomethacin, 0_placebo    0.494      0.125 -4.04   <0.001 0.248   0.74
##
## Columns: term, contrast, estimate, std.error, statistic, p.value, conf.low, conf.high, predicted, pr
```

### 2

```
fit_ad <- glm(outcome~rx+amp+pep, study_data, family = binomial)
avg_comparisons(model = fit_ad,variables = "rx",vcov = "HC3",
                comparison =  \(hi, lo) (mean(hi)/(1 - mean(hi)))/(mean(lo)/(1 - mean(lo))) ,
                hypothesis = 1)
```

```
##
##  Term                  Contrast Estimate Std. Error    z Pr(>|z|) 2.5 % 97.5 %
##    rx 1_indomethacin, 0_placebo    0.527      0.121 -3.9   <0.001 0.288  0.765
##
## Columns: term, contrast, estimate, std.error, statistic, p.value, conf.low, conf.high, predicted, pr
```

### 3

```
OR <- exp(as.numeric(coef(fit_ad)[2]))
OR
```

```
## [1] 0.4815847
```

```
exp(coefci(fit_ad, vcov. = vcovHC))
```

```
##                       2.5 %     97.5 %
## (Intercept)       0.1132212  0.2203739
## rx1_indomethacin 0.2882002  0.8047315
## amp1_yes          1.3813463 11.3547572
## pep1_yes          1.5189717  4.6514065
```

**4**

- 1 and 2

**5**

- 1 and 2

**P2**

**1.**

```
mu_A <- \(A,X) plogis(X^2 + A)
DGP_A <- function(n){
  tibble(
    X = runif(n),
    A = rbern(n, prob = 0.5),
    Y1 = rbern(n, prob = mu_A(1,X)),
    Y0 = rbern(n, prob = mu_A(0,X)),
    Y = ifelse(A,Y1,Y0)
  )
}

A <- DGP_A(10000)
fit_a <- glm(Y~A+X, A, family = binomial)

RR_TRUEA <- A %$% {
  mean(Y1)/mean(Y0)
}

mu_B <- \(A,X) plogis(X^2 + A + X*A)
DGP_B <- function(n){
  tibble(
    X = runif(n),
    A = rbern(n, prob = 0.5),
    Y1 = rbern(n, prob = mu_B(1,X)),
    Y0 = rbern(n, prob = mu_B(0,X)),
    Y = ifelse(A,Y1,Y0)
  )
}

B <- DGP_B(10000)
```

```
RR_TRUEB <- B %$% {
  mean(Y1)/mean(Y0)
}
fit_b <- glm(Y~A+X, B, family = binomial)
##############################
A_EST <- avg_comparisons(model = fit_a,variables = "A",vcov = "HC3",
                comparison =  "ratioavg" ,
                hypothesis = 1)

A_EST
```

```
##
##  Term          Contrast Estimate Std. Error    z Pr(>|z|) 2.5 % 97.5 %
##     A mean(1) / mean(0)     1.35      0.019 18.6   <0.001  1.32   1.39
##
## Columns: term, contrast, estimate, std.error, statistic, p.value, conf.low, conf.high, predicted, pr
```

```
B_EST <- avg_comparisons(model = fit_b,variables = "A",vcov = "HC3",
                comparison =  "ratioavg" ,
                hypothesis = 1)

B_EST
```

```
##
##  Term          Contrast Estimate Std. Error    z Pr(>|z|) 2.5 % 97.5 %
##     A mean(1) / mean(0)     1.46     0.0194 23.9   <0.001  1.43    1.5
##
## Columns: term, contrast, estimate, std.error, statistic, p.value, conf.low, conf.high, predicted, pr
```

```
tibble(A_TRUE = RR_TRUEA, A_estimate = as.numeric(A_EST[3]), B_TRUE = RR_TRUEB, B_estimate = as.numeric
```

```
## # A tibble: 1 x 4
##   A_TRUE A_estimate B_TRUE B_estimate
##    <dbl>      <dbl>  <dbl>      <dbl>
## 1   1.35       1.35   1.45       1.46
```

**2.**

```
mu_C <- \(A,X) plogis(X + A)
DGP_C <- function(n){
  tibble(
    X = runif(n),
    A = rbern(n, prob = 0.5),
    Y1 = rbern(n, prob = mu_C(1,X)),
    Y0 = rbern(n, prob = mu_C(0,X)),
    Y = ifelse(A,Y1,Y0)
  )
}
```

```r
mu_D <- \(A,X) plogis(X + A)
DGP_D <- function(n){
  tibble(
    X = runif(n),
    A = rbern(n, prob = plogis(X^2)),
    Y1 = rbern(n, prob = mu_D(1,X)),
    Y0 = rbern(n, prob = mu_D(0,X)),
    Y = ifelse(A,Y1,Y0)
  )
}
#### sampling distributions##############
samp_C <- map_df(1:1000, function(.x) {
  C <- DGP_C(10000)
  fit_c <- glm(Y~A+X, C, family = binomial)
  Est <- avg_comparisons(model = fit_c,variables = "A",vcov = "HC3",
                comparison =  "ratioavg" ,
                hypothesis = 1)
  distb <- tibble(estimate = as.numeric(Est[3]))
  return(distb)
})
samp_C_sd <- sd(samp_C$estimate)
samp_D <- map_df(1:1000, function(.x) {
  D <- DGP_D(10000)
  fit_d <- glm(Y~A+X, D, family = binomial)
  Est <- avg_comparisons(model = fit_d,variables = "A",vcov = "HC3",
                comparison =  "ratioavg" ,
                hypothesis = 1)
  distb <- tibble(estimate = as.numeric(Est[3]))
  return(distb)
})
samp_D_sd <- sd(samp_D$estimate)
#### individual estimates#################
C <- DGP_C(10000)
fit_c <- glm(Y~A+X, C, family = binomial)
EST_C <- avg_comparisons(model = fit_c,variables = "A",vcov = "HC3",
                comparison =  "ratioavg" ,
                hypothesis = 1)
EST_CSE <- as.numeric(EST_C[4])



D <- DGP_D(10000)
fit_d <- glm(Y~A+X, D, family = binomial)
EST_D <- avg_comparisons(model = fit_d,variables = "A",vcov = "HC3",
                comparison =  "ratioavg" ,
                hypothesis = 1)
EST_DSE <- as.numeric(EST_D[4])

### solution ###############

solution <- tibble(Robust_SE_C = EST_CSE, true_samp_sd_c = samp_C_sd, Robust_SE_D = EST_DSE, true_samp_
solution
```

```
## # A tibble: 1 x 4
##   Robust_SE_C true_samp_sd_c Robust_SE_D true_samp_sd_d
##         <dbl>          <dbl>       <dbl>          <dbl>
## 1      0.0168         0.0163      0.0185         0.0184
```