emmeans and varpred comparison

Bicko, Jonathan & Ben

2021 Oct 06 (Wed)

Continuous predictors

No interaction

```
## [1] "Truth:"
```

```
## [1] 1.235385
```

```
##     model      fit
## 1 emmeans 1.235385
## 2 varpred 1.235385
```

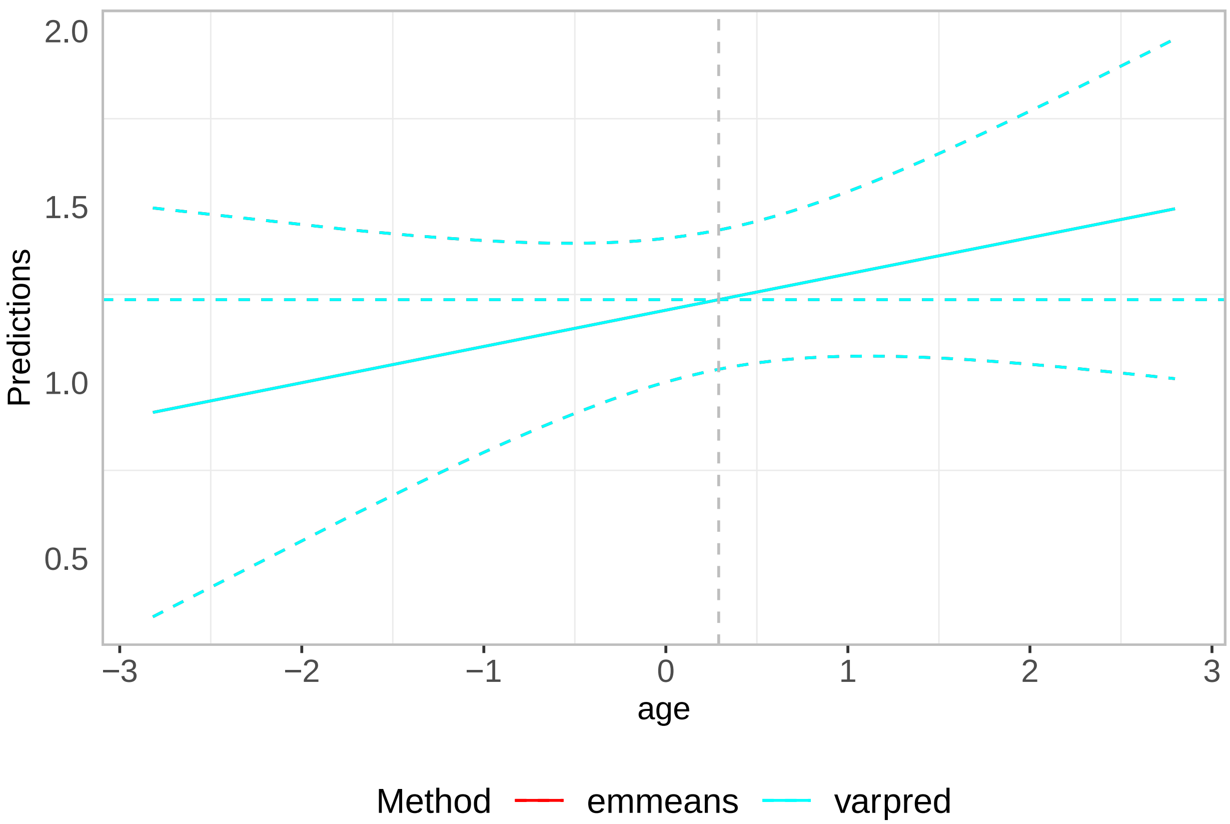

Figure 1: Continuous: no interaction

Interaction between non-focal predictors

```
## [1] "Truth:"
```

```
## [1] 1.750468

##     model      fit
## 1 emmeans 1.782517
## 2 varpred 1.750468
```

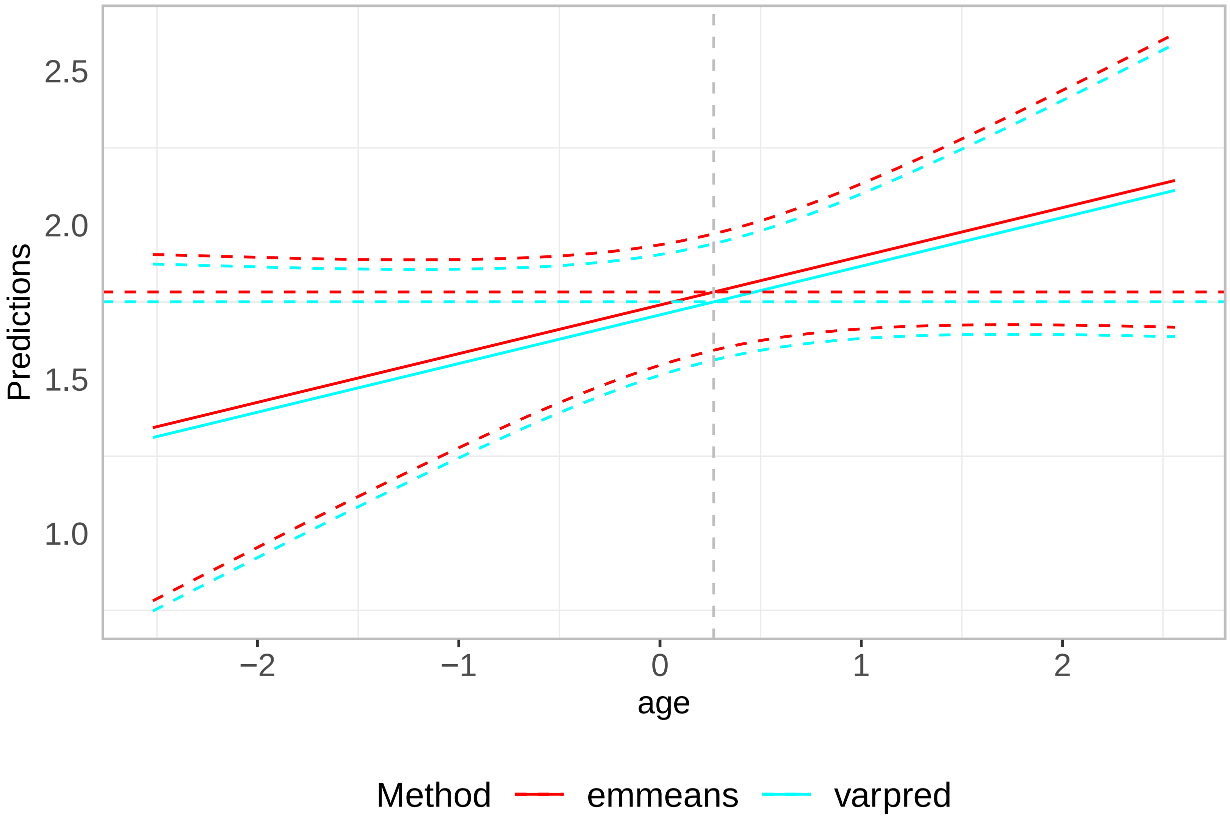

Figure 2: Continuous: non-focal predictors interaction

Interaction between focal and non-focal predictors

```
## [1] "Truth:"

## [1] 1.590765

##     model      fit
## 1 emmeans 1.590369
## 2 varpred 1.590765
```

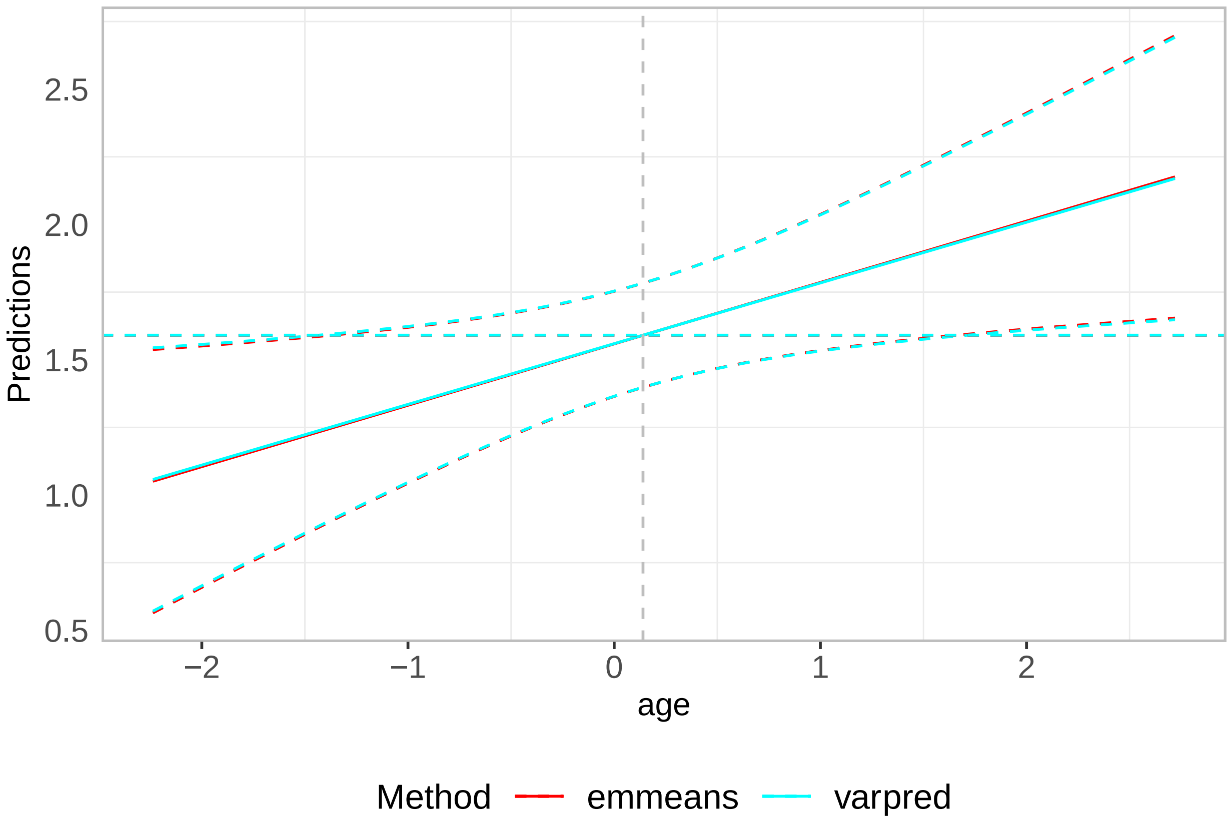

Figure 3: Continuous: interaction betweem focal and non-focal predictors

Categorical predictors

No interaction

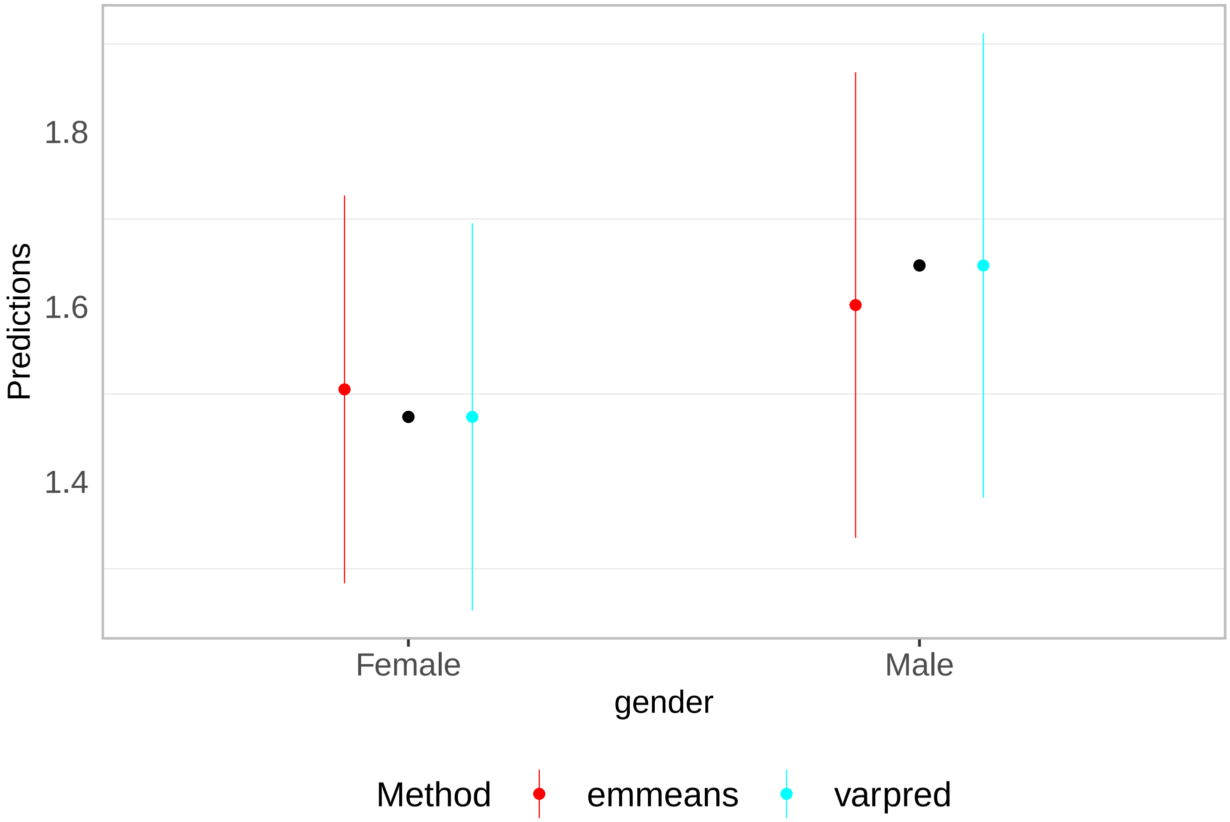

Figure 4: Categorical: no interaction

Why the differences?

We actually don't know which method is correct. Let us consider a simple simulation:

- Outcome: hhsize
- Predictors:
  - Gender (40% Males)
  - Age: Females are slightly older

```
## Simulation
N <- 100
extraAge <- 0.4
meanAge <- 0.5
sdAge <- 1
prop <- c(0.6, 0.4)
betas <- c(1.5, 0.5, 0.5)
gender <- sample(c("Female", "Male"), N, replace = TRUE, prob = prop)
df <- (data.frame(gender=gender)
    %>% mutate(muAge=ifelse(gender=="Female", meanAge+extraAge, meanAge)
        , age=rnorm(N, muAge, sdAge)
    )
    %>% select(-muAge)
```

```r
)

mm <- model.matrix(~gender+age, df)
df$hhsize <- rnorm(N, mean=as.vector(mm %*% betas), sd=1)

## Model
mod <- lm(hhsize~gender+age, data=df)
print(mod)
```

```
##
## Call:
## lm(formula = hhsize ~ gender + age, data = df)
##
## Coefficients:
## (Intercept)   genderMale          age
##      1.2987       0.7385       0.4739
```

Let us take a look at the observed marginals:

```r
observed_margins <- (df
    %>% group_by(gender)
    %>% summarise_all(mean)
)
observed_margins
```

```
## # A tibble: 2 x 3
##   gender   age hhsize
##   <chr>  <dbl>  <dbl>
## 1 Female 0.897   1.72
## 2 Male   0.647   2.34
```

- **varpred** constructs model matrix by averaging *age* within the levels of *gender*. The population average is the weighted (by the observed proportions) average of these averages:

```r
varpred_pred <- varpred(mod, "gender", within.category=TRUE, returnall=TRUE)
varpred_mm <- varpred_pred$raw$model.matrix
print(varpred_mm)
```

```
##        (Intercept) genderMale       age
## Female           1          0 0.8969018
## Male             1          1 0.6473886
```

Estimates:

```
## Call:
## varpred(mod = mod, focal_predictors = "gender", within.category = TRUE,
##     returnall = TRUE)
##
##   gender      fit        se      lwr      upr
## 1 Female 1.723710 0.1351851 1.455405 1.992015
## 2   Male 2.343957 0.1407051 2.064697 2.623218
```

By hand calculation:

```r
varpred_mm %*% coef(mod)
```

```
##            [,1]
## Female 1.723710
```

```
## Male   2.343957
```

- **emmeans** uses the population average and seems not apply the weights:

```r
emmeans_grid <- ref_grid(mod)
emmeans_mm <- emmeans_grid@grid
print(emmeans_mm)
```

```
##   gender       age .wgt.
## 1 Female 0.7771354    52
## 2   Male 0.7771354    48
```

Estimates:

```
##  gender emmean    SE df lower.CL upper.CL
##  Female   1.67 0.136 97     1.40     1.94
##  Male     2.41 0.141 97     2.13     2.69
##
## Confidence level used: 0.95
```

By hand calculation:

```r
emmeans_mm <- emmeans_pred@linfct
print(emmeans_mm)
```

```
##      (Intercept) genderMale       age
## [1,]           1          0 0.7771354
## [2,]           1          1 0.7771354
```

```r
as.matrix(emmeans_mm) %*% coef(mod)
```

```
##          [,1]
## [1,] 1.666952
## [2,] 2.405445
```

Mathematically, in **varpred**

$$\mathbb{E}(Y|X) = \begin{cases} (\beta_0 + \alpha) + \beta_1 \bar{Age}_M, & \text{if } gender = Male \\ \\ \beta_0 + \beta_1 \bar{Age}_F, & \text{if } gender = Female \end{cases} \tag{1}$$

emmeans

$$\mathbb{E}(Y|X) = \begin{cases} (\beta_0 + \alpha) + \beta_1 \bar{Age}, & \text{if } gender = Male \\ \\ \beta_0 + \beta_1 \bar{Age}, & \text{if } gender = Female \end{cases} \tag{2}$$