Estimating marginal means using the regression equation

Nick Michalak

October 28th, 2017

packages

```
library(tidyverse)
library(haven)
library(multcomp)
```

data

dummy 1

```
dummy1 <- c(0, 1) %>% rep(each = 50)
```

contrast 1

```
cont1 <- (c(-1, 1) / 2) %>% rep(each = 50)
```

dummy 2

```
dummy2 <- c(0, 1) %>% rep(each = 25, times = 2)
```

contrast 2

```
cont2 <- (c(-1, 1) / 2) %>% rep(each = 25, times = 2)
```

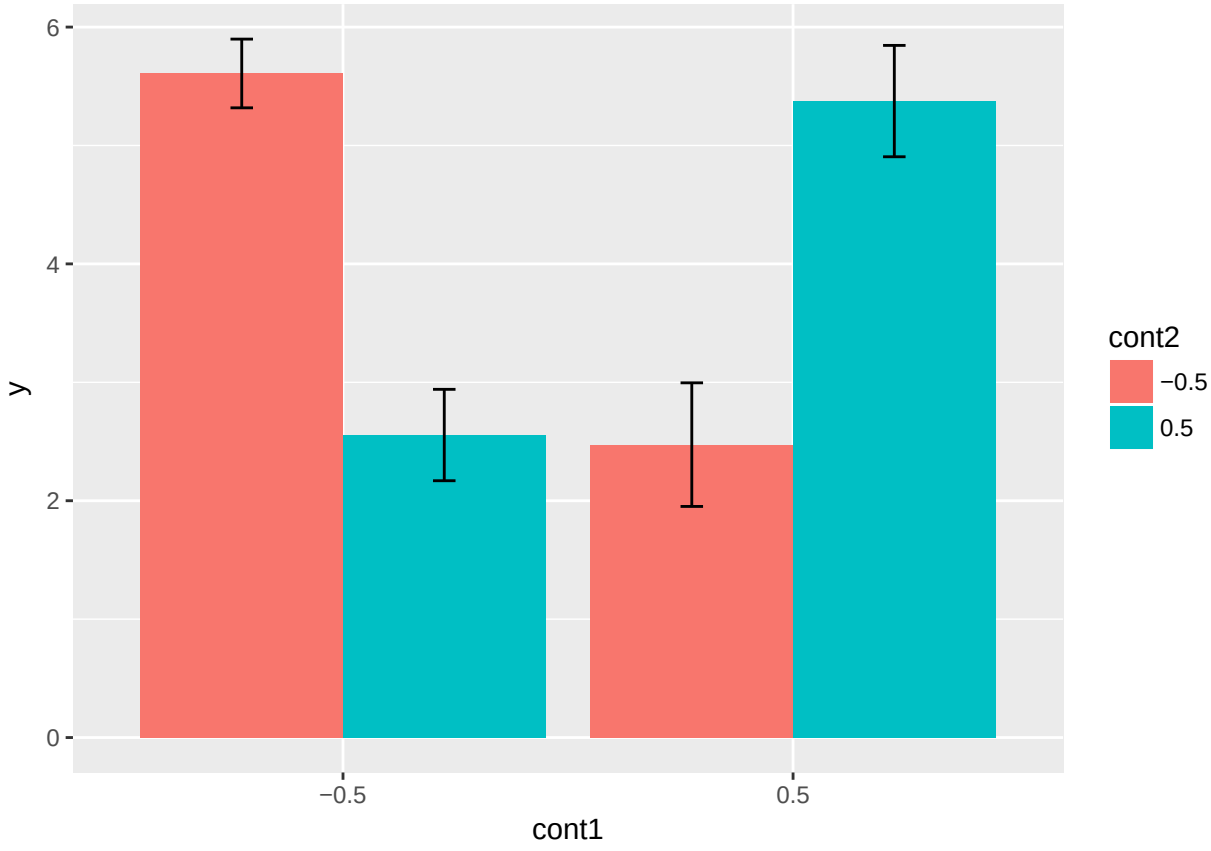
y

```
# set randomizer seed so that results are reproducible
set.seed(100)

y <- c(5.5, 2.5, 2.5, 5.5) %>% map(function(m) rnorm(n = 25, mean = m, sd = 1)) %>% unlist()
```

plot

```r
tibble(y, cont1, cont2) %>%
  mutate(cont1 = cont1 %>% factor(),
         cont2 = cont2 %>% factor()) %>%
  ggplot(mapping = aes(x = cont1, y = y, fill = cont2)) +
  stat_summary(geom = "bar", fun.data = mean_cl_normal, position = position_dodge(0.9)) +
  stat_summary(geom = "errorbar", fun.data = mean_cl_normal, position = position_dodge(0.9), width = 0.
```



cell means

```r
tibble(dummy1, dummy2, y) %>%
  group_by(dummy1, dummy2) %>%
  summarize(Mean = mean(y),
            SD = sd(y),
            n = sum(!is.na(y)))
```

```
## # A tibble: 4 x 5
## # Groups:   dummy1 [?]
##   dummy1 dummy2     Mean        SD     n
##    <dbl>  <dbl>    <dbl>     <dbl> <int>
## 1      0      0 5.608172 0.7028586    25
## 2      0      1 2.554846 0.9348886    25
## 3      1      0 2.473703 1.2646553    25
## 4      1      1 5.374929 1.1392753    25
```

models

dummy model

```
lm_dummy <- lm(y ~ dummy1 * dummy2)
```

summary

```
lm_dummy %>% summary()
```

```
##
## Call:
## lm(formula = y ~ dummy1 * dummy2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.24563 -0.68754 -0.02556  0.61269  2.60826
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     5.6082     0.2065   27.16   <2e-16 ***
## dummy1         -3.1345     0.2921  -10.73   <2e-16 ***
## dummy2         -3.0533     0.2921  -10.45   <2e-16 ***
## dummy1:dummy2   5.9546     0.4131   14.42   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.033 on 96 degrees of freedom
## Multiple R-squared:  0.6848, Adjusted R-squared:  0.6749
## F-statistic: 69.51 on 3 and 96 DF,  p-value: < 2.2e-16
```

the second and third cofficient, dummy1 and dummy2, are simple effects. dummy1 is the difference between the two red bars or the effect of a 1 unit change in dummy1 when dummy2 = 0; dummy2 is the difference between the far left red bar and the blue bar next to it or the effect of a 1 unite change in dummy2 when dummy1 = 0.

use regression equation to get marginal means from dummy codes

the first coefficient – the intercept – gives the marginal mean for a specific cell when you weigh the regression cofficients using the values that represent that cell for that cofficient's predictor variable. In all the coefficient weights below, the intercept always gets weight = 1 because that's the coefficient we're interested in.

far left bar

the far left bar's mean is the value of the intercept when dummy1 = 0 and dummy2 = 0, so the first cofficient = 0, the second = 0, and the third = 0 * 0 = 0.

```
lm_dummy %>%
  glht(linfct = rbind(c(1, 0, 0, 0))) %>%
  summary(test = adjusted("none"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = y ~ dummy1 * dummy2)
##
## Linear Hypotheses:
##        Estimate Std. Error t value Pr(>|t|)
## 1 == 0   5.6082     0.2065   27.16   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- none method)
```

second bar from left bar

the second bar from the left's mean is the value of the intercept when dummy1 = 0 and dummy2 = 1, so the first cofficient = 0, the second = 0, and the third = 0 * 1 = 0.

```
lm_dummy %>%
  glht(linfct = rbind(c(1, 0, 1, 0))) %>%
  summary(test = adjusted("none"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = y ~ dummy1 * dummy2)
##
## Linear Hypotheses:
##        Estimate Std. Error t value Pr(>|t|)
## 1 == 0   2.5548     0.2065   12.37   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- none method)
```

third bar from left

the third bar from the left's mean is the value of the intercept when dummy1 = 1 and dummy2 = 0, so the first cofficient = 1, the second = 0, and the third = 1 * 0 = 0.

```
lm_dummy %>%
  glht(linfct = rbind(c(1, 1, 0, 0))) %>%
  summary(test = adjusted("none"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = y ~ dummy1 * dummy2)
##
## Linear Hypotheses:
##        Estimate Std. Error t value Pr(>|t|)
## 1 == 0   2.4737     0.2065   11.98   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- none method)
```

fourth bar from left (the far right bar)

the fourth bar from the left's mean (the far right bar) is the value of the intercept when dummy1 = 1 and dummy2 = 1, so the first cofficient = 1, the second = 1, and the third = 1 * 1 = 1.

```
lm_dummy %>%
  glht(linfct = rbind(c(1, 1, 1, 1))) %>%
  summary(test = adjusted("none"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = y ~ dummy1 * dummy2)
##
## Linear Hypotheses:
##        Estimate Std. Error t value Pr(>|t|)
## 1 == 0   5.3749     0.2065   26.02   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- none method)
```

contrast model

```
lm_cont <- lm(y ~ cont1 * cont2)
```

summary

```
lm_cont %>% summary()
```

```
##
## Call:
## lm(formula = y ~ cont1 * cont2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.24563 -0.68754 -0.02556  0.61269  2.60826
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.00291    0.10326  38.764   <2e-16 ***
## cont1       -0.15719    0.20653  -0.761    0.448
## cont2       -0.07605    0.20653  -0.368    0.714
## cont1:cont2  5.95455    0.41305  14.416   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.033 on 96 degrees of freedom
## Multiple R-squared:  0.6848, Adjusted R-squared:  0.6749
## F-statistic: 69.51 on 3 and 96 DF,  p-value: < 2.2e-16
```

the second and third cofficient, cont1 and cont2, are main effects. cont1 is the difference between the average of the left bars and the average of the right bars or the effect of a 1 unit change in

cont1 when cont2 = 0. cont2 is the difference between the average of the red bars and the average of the blue bars or the effect of a 1 unit change in cont2 when cont1 = 0.

use regression equation to get marginal means from contrast codes

far left bar

same as with the dummy codes but we're using difefrent weights: 1 for the intercept, cont1 = -0.5, cont2 = -0.5, and their product = -0.5 * -0.5 = 0.25

```
lm_cont %>%
  glht(linfct = rbind(c(1, -0.5, -0.5, 0.25))) %>%
  summary(test = adjusted("none"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = y ~ cont1 * cont2)
##
## Linear Hypotheses:
##        Estimate Std. Error t value Pr(>|t|)
## 1 == 0   5.6082     0.2065   27.16   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- none method)
```

second bar from left bar

1 for the intercept, cont1 = -0.5, cont2 = 0.5, and their product = -0.5 * 0.5 = -0.25

```
lm_cont %>%
  glht(linfct = rbind(c(1, -0.5, 0.5, -0.25))) %>%
  summary(test = adjusted("none"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = y ~ cont1 * cont2)
##
## Linear Hypotheses:
##        Estimate Std. Error t value Pr(>|t|)
## 1 == 0   2.5548     0.2065   12.37   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- none method)
```

third bar from left

1 for the intercept, cont1 = 0.5, cont2 = -0.5, and their product = 0.5 * -0.5 = -0.25

```
lm_cont %>%
  glht(linfct = rbind(c(1, 0.5, -0.5, -0.25))) %>%
  summary(test = adjusted("none"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
```

```
##
## Fit: lm(formula = y ~ cont1 * cont2)
##
## Linear Hypotheses:
##        Estimate Std. Error t value Pr(>|t|)
## 1 == 0   2.4737     0.2065   11.98   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- none method)
```

fourth bar from left (the far right bar)

1 for the intercept, cont1 = 0.5, cont2 = 0.5, and their product = 0.5 * 0.5 = 0.25

```
lm_cont %>%
  glht(linfct = rbind(c(1, 0.5, 0.5, 0.25))) %>%
  summary(test = adjusted("none"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = y ~ cont1 * cont2)
##
## Linear Hypotheses:
##        Estimate Std. Error t value Pr(>|t|)
## 1 == 0   5.3749     0.2065   26.02   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- none method)
```

export data

```
tibble(y, cont1, cont2, dummy1, dummy2) %>%
  write_sav(path = "~/Desktop/novum_R_ganum/dummy_cont_data.sav")
```