Contrasts

Sofía Fernández

October 18, 2018

Loading the data

The data was downloaded from the book's resources and loaded as a comma separated values file. Afterwards the dataset was filtered so only the woman without diabetes were kept.

```
hers <- read_csv("hersdata.csv")
hers_nodi <- filter(hers, diabetes == "no")
```

```
## # A tibble: 2,032 x 37
##    HT      age raceth nonwhite smoking drinkany exercise physact globrat
##    <chr> <int> <chr>  <chr>    <chr>   <chr>    <chr>    <chr>   <chr>
##  1 plac~    70 Afric~ yes      no      no       no       much m~ good
##  2 plac~    62 Afric~ yes      no      no       no       much l~ good
##  3 plac~    64 White  no       yes     yes      no       much l~ good
##  4 plac~    65 White  no       no      no       no       somewh~ good
##  5 horm~    68 Afric~ yes      no      yes      no       about ~ good
##  6 horm~    69 White  no       no      no       yes      much m~ very g~
##  7 horm~    61 White  no       no      yes      yes      about ~ very g~
##  8 horm~    62 White  no       yes     yes      no       somewh~ good
##  9 plac~    72 Afric~ yes      no      no       no       about ~ fair
## 10 horm~    73 White  no       no      no       no       somewh~ good
## # ... with 2,022 more rows, and 28 more variables: poorfair <chr>,
## #   medcond <int>, htnmeds <chr>, statins <chr>, diabetes <chr>,
## #   dmpills <chr>, insulin <chr>, weight <dbl>, BMI <dbl>, waist <dbl>,
## #   WHR <dbl>, glucose <int>, weight1 <dbl>, BMI1 <dbl>, waist1 <dbl>,
## #   WHR1 <dbl>, glucose1 <int>, tchol <int>, LDL <dbl>, HDL <int>,
## #   TG <int>, tchol1 <int>, LDL1 <dbl>, HDL1 <int>, TG1 <int>, SBP <int>,
## #   DBP <int>, age10 <dbl>
```

Ordering the Physical activity (`physact`)

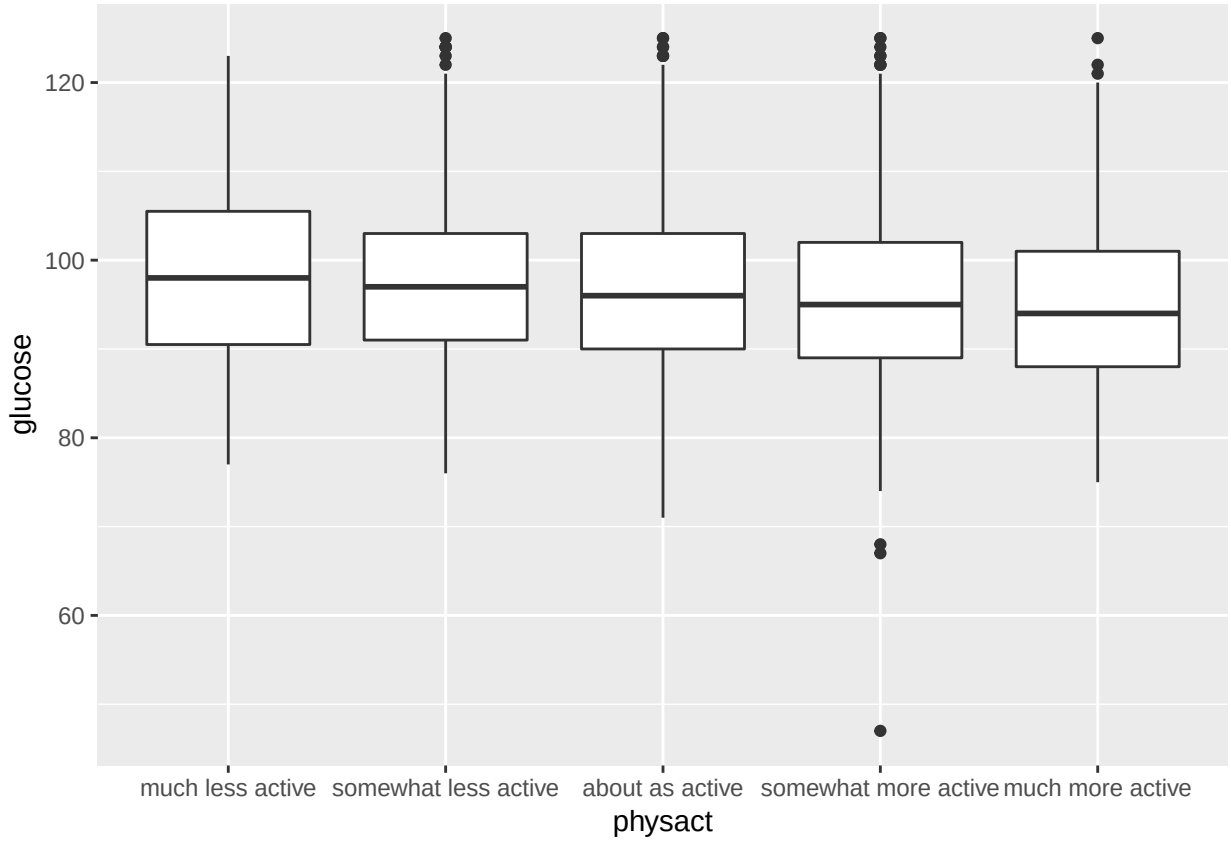
```
physact_levels <- c("much less active",
                    "somewhat less active",
                    "about as active",
                    "somewhat more active",
                    "much more active")

hers_nodi <- mutate(hers_nodi,
                    physact = ordered(physact,
                                      levels = physact_levels))
```

```
## [1] much more active     much less active     much less active
## [4] somewhat less active about as active      much more active
## 5 Levels: much less active < somewhat less active < ... < much more active
```

Plot

```
ggplot(hers_nodi, aes(physact, glucose)) +
    geom_boxplot(na.rm = TRUE)
```



LM

First the linear model was created and saved as `glucose_fit_act`. Seeing if the glucose variable changes can be explained by the changes in the predictor variable of physical activity.

```
glucose_fit_act <- lm(glucose ~ physact, data = hers_nodi)
```

ANOVA

```
## Anova Table (Type II tests)
##
## Response: glucose
##            Sum Sq   Df F value   Pr(>F)
## physact     1673    4   4.431 0.001441 **
## Residuals 191345 2027
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

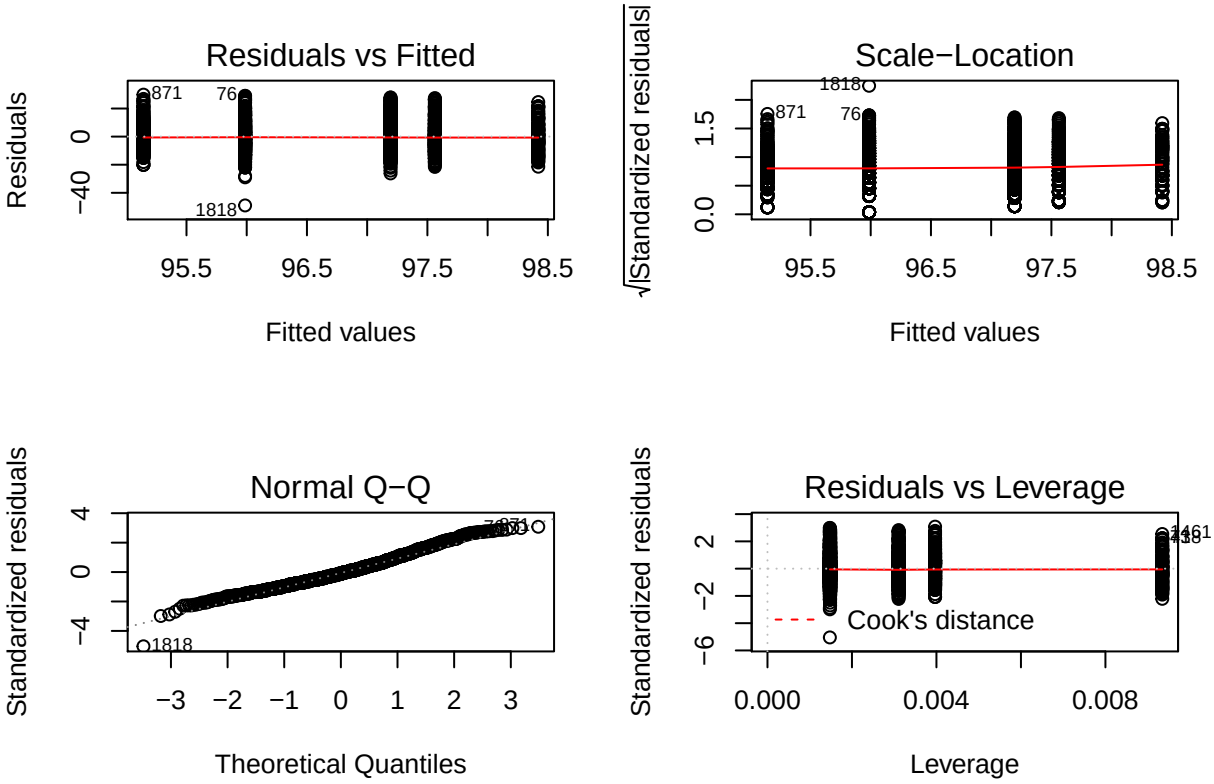
The Analysis of Variance shows the model to be significant, $ F_{(4)} = 4.431, p < 0.01 $.

Summary

```
##
## Call:
## lm(formula = glucose ~ physact, data = hers_nodi)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -48.987  -6.987  -0.987   5.806  29.857
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 96.86132    0.27052 358.061  < 2e-16 ***
## physact.L   -2.57119    0.73891  -3.480 0.000513 ***
## physact.Q   -0.21667    0.65575  -0.330 0.741119
## physact.C   -0.04013    0.54655  -0.073 0.941481
## physact^4    0.30291    0.43458   0.697 0.485876
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.716 on 2027 degrees of freedom
## Multiple R-squared:  0.008668,   Adjusted R-squared:  0.006712
## F-statistic: 4.431 on 4 and 2027 DF,  p-value: 0.001441
```

The summary of the model, because the `physact` variable is factored as ordered, instead of giving the differences to a basal value of it, it uses orthogonal polynomial contrasts and shows that the values of the variable behave as a linear equation.

Plots



Contrasts

EMMEANS

With the package `emmeans` and the homonymous function an object of the estimated marginal means (EMMs). Later the contrasts were created and saved in an object to be used later. The contrasts were applied following the order of the `physact` variable. Using these contrasts the differences were explored with a "sidak" method. It makes adjustments as if the estimates were independent.

```
glucose_lstsqr <- emmeans(glucose_fit_act, "physact")
```

```
##  physact                emmean        SE   df lower.CL  upper.CL
##  much less active     98.42056 0.9392676 2027 96.57853 100.26259
##  somewhat less active 97.56211 0.5414437 2027 96.50027  98.62396
##  about as active      97.19436 0.3742409 2027 96.46043  97.92830
##  somewhat more active 95.98671 0.3734108 2027 95.25440  96.71902
##  much more active     95.14286 0.6120416 2027 93.94256  96.34315
##
## Confidence level used: 0.95
```

```
# Contrasts
Contrasts_glu = list(MAvsLA          = c(-1, -1,  0,  1,  1),
                     InteractMuchSo  = c( 1, -1,  0, -1,  1),
                     MAvsLAforMuch   = c(-1,  0,  0,  0,  1),
                     MAvsLAforSome   = c( 0, -1,  0,  1,  0),
                     phyActvsControl = c( 1,  1, -4,  1,  1),
```

```
                    MLAvsC          = c( 1,  0, -1,  0,  0),
                    MMAvsC          = c( 0,  0, -1,  0,  1),
                    SLAvsC          = c( 0,  1, -1,  0,  0),
                    SMAvsC          = c( 0,  0, -1,  1,  0))
```

Sidak

Using these contrasts the differences were explored with a "sidak" method. It makes adjustments as if the estimates were independent.

```
contrast(glucose_lstsqr, Contrasts_glu, adjust="sidak")
```

```
##  contrast           estimate        SE   df t.ratio p.value
##  MAvsLA          -4.85310935 1.2997752 2027  -3.734  0.0017
##  InteractMuchSo   0.01460003 1.2997752 2027   0.011  1.0000
##  MAvsLAforMuch   -3.27770360 1.1210792 2027  -2.924  0.0310
##  MAvsLAforSome   -1.57540575 0.6577210 2027  -2.395  0.1406
##  phyActvsControl -1.66521232 1.9825025 2027  -0.840  0.9901
##  MLAvsC           1.22619873 1.0110786 2027   1.213  0.8996
##  MMAvsC          -2.05150487 0.7173920 2027  -2.860  0.0379
##  SLAvsC           0.36774978 0.6581926 2027   0.559  0.9996
##  SMAvsC          -1.20765596 0.5286700 2027  -2.284  0.1849
##
## P value adjustment: sidak method for 9 tests
```

The differences were significant for the comparisons between *more v less* activity, the extreme levels of *much more v much less* activity, and *much more v control* with a $ p_{corr} < 0.05 $

Dunnettx

For comparison the same contrasts were compared with the `dunnettx` method. It uses `emmeans`'s own ad hoc approximation to the Dunnett distribution for a family of esti- mates having pairwise correlations of 0.5.

```
contrast(glucose_lstsqr, Contrasts_glu, adjust="dunnettx")
```

```
##  contrast           estimate        SE   df t.ratio p.value
##  MAvsLA          -4.85310935 1.2997752 2027  -3.734  0.0017
##  InteractMuchSo   0.01460003 1.2997752 2027   0.011  1.0000
##  MAvsLAforMuch   -3.27770360 1.1210792 2027  -2.924  0.0268
##  MAvsLAforSome   -1.57540575 0.6577210 2027  -2.395  0.1115
##  phyActvsControl -1.66521232 1.9825025 2027  -0.840  0.9134
##  MLAvsC           1.22619873 1.0110786 2027   1.213  0.7392
##  MMAvsC          -2.05150487 0.7173920 2027  -2.860  0.0324
##  SLAvsC           0.36774978 0.6581926 2027   0.559  0.9782
##  SMAvsC          -1.20765596 0.5286700 2027  -2.284  0.1441
##
## P value adjustment: dunnettx method for 9 tests
```

The same differences were significant with just little variation in the values of p.

None

Later the same contrasts were explored without any method of adustment of multiple comparisons.

```
contrast(glucose_lstsqr, Contrasts_glu, adjust="none")
```

```
##  contrast           estimate        SE   df t.ratio p.value
##  MAvsLA          -4.85310935 1.2997752 2027  -3.734  0.0002
##  InteractMuchSo   0.01460003 1.2997752 2027   0.011  0.9910
##  MAvsLAforMuch   -3.27770360 1.1210792 2027  -2.924  0.0035
##  MAvsLAforSome   -1.57540575 0.6577210 2027  -2.395  0.0167
##  phyActvsControl -1.66521232 1.9825025 2027  -0.840  0.4010
##  MLAvsC           1.22619873 1.0110786 2027   1.213  0.2254
##  MMAvsC          -2.05150487 0.7173920 2027  -2.860  0.0043
##  SLAvsC           0.36774978 0.6581926 2027   0.559  0.5764
##  SMAvsC          -1.20765596 0.5286700 2027  -2.284  0.0225
```

Without any method of correction, the same comparisons as before are shown as significant with lower p values, and also the comparisons between the levels of *somewhat more v somewhat less* and *somewhat more v control*.

Multcomp

The same process was followed but with the package of `multcomp`. First the contrasts were saved as an object to be used later.

```
Input = ("
Contrast.Name     MLA   SLA  AAA  SMA  MMA
 MAvsLA            -1    -1    0    1    1
 InteractMuchSo     1    -1    0   -1    1
 MAvsLAforMuch     -1     0    0    0    1
 MAvsLAforSome      0    -1    0    1    0
 phyActvsControl    1     1   -4    1    1
 MLAvsC             1     0   -1    0    0
 MMAvsC             0     0   -1    0    1
 SLAvsC             0     1   -1    0    0
 SMAvsC             0     0   -1    1    0
")


Cont_glucose_Matriz = as.matrix(read.table(textConnection(Input), header=TRUE, row.names=1))
```

```
##                  MLA SLA AAA SMA MMA
## MAvsLA            -1  -1   0   1   1
## InteractMuchSo     1  -1   0  -1   1
## MAvsLAforMuch     -1   0   0   0   1
## MAvsLAforSome      0  -1   0   1   0
## phyActvsControl    1   1  -4   1   1
## MLAvsC             1   0  -1   0   0
## MMAvsC             0   0  -1   0   1
## SLAvsC             0   1  -1   0   0
## SMAvsC             0   0  -1   1   0
```

General Linear Hyphoteses

In an object named `G` the general linear hypotheses containing the multiple comparisons of the contrasts was saved with the `mcp` function on in the `linfct` argument of the function.

```
G = glht(glucose_fit_act, linfct = mcp(physact = Cont_glucose_Matriz))
```

Single-step

Using an adjustment of *single-step* the comparisons were explored.

```
summary(G, test=adjusted("single-step"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: User-defined Contrasts
##
##
## Fit: lm(formula = glucose ~ physact, data = hers_nodi)
##
## Linear Hypotheses:
##                       Estimate Std. Error t value Pr(>|t|)
## MAvsLA == 0           -4.8531     1.2998  -3.734  0.00157 **
## InteractMuchSo == 0    0.0146     1.2998   0.011  1.00000
## MAvsLAforMuch == 0    -3.2777     1.1211  -2.924  0.02470 *
## MAvsLAforSome == 0    -1.5754     0.6577  -2.395  0.10219
## phyActvsControl == 0  -1.6652     1.9825  -0.840  0.90595
## MLAvsC == 0            1.2262     1.0111   1.213  0.71635
## MMAvsC == 0           -2.0515     0.7174  -2.860  0.02971 *
## SLAvsC == 0            0.3678     0.6582   0.559  0.97753
## SMAvsC == 0           -1.2077     0.5287  -2.284  0.13239
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Same as with the other package, the same levels of comparisons show to be significant with just little variation of p.

Bonferroni & FDR

The same comparisons were adjusted with the bonferroni and FDR methods for comparison.

```
summary(G, test=adjusted("bonferroni"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: User-defined Contrasts
##
##
## Fit: lm(formula = glucose ~ physact, data = hers_nodi)
##
## Linear Hypotheses:
##                       Estimate Std. Error t value Pr(>|t|)
## MAvsLA == 0           -4.8531     1.2998  -3.734  0.00174 **
## InteractMuchSo == 0    0.0146     1.2998   0.011  1.00000
## MAvsLAforMuch == 0    -3.2777     1.1211  -2.924  0.03148 *
## MAvsLAforSome == 0    -1.5754     0.6577  -2.395  0.15029
```

```
## phyActvsControl == 0  -1.6652     1.9825  -0.840  1.00000
## MLAvsC == 0            1.2262     1.0111   1.213  1.00000
## MMAvsC == 0           -2.0515     0.7174  -2.860  0.03856 *
## SLAvsC == 0            0.3678     0.6582   0.559  1.00000
## SMAvsC == 0           -1.2077     0.5287  -2.284  0.20210
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- bonferroni method)
```

The same comparisons show to be significant.

```
summary(G, test=adjusted("fdr"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: User-defined Contrasts
##
##
## Fit: lm(formula = glucose ~ physact, data = hers_nodi)
##
## Linear Hypotheses:
##                       Estimate Std. Error t value Pr(>|t|)
## MAvsLA == 0           -4.8531     1.2998  -3.734  0.00174 **
## InteractMuchSo == 0    0.0146     1.2998   0.011  0.99104
## MAvsLAforMuch == 0    -3.2777     1.1211  -2.924  0.01285 *
## MAvsLAforSome == 0    -1.5754     0.6577  -2.395  0.03757 *
## phyActvsControl == 0  -1.6652     1.9825  -0.840  0.51561
## MLAvsC == 0            1.2262     1.0111   1.213  0.33804
## MMAvsC == 0           -2.0515     0.7174  -2.860  0.01285 *
## SLAvsC == 0            0.3678     0.6582   0.559  0.64846
## SMAvsC == 0           -1.2077     0.5287  -2.284  0.04042 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- fdr method)
```

Being less strict than the previous methods, FDR show, besides the other significant comparisons, significant differences in the *somewhat more v somewhat less*, and *somewhat more v control* levels of comparisons.