MANOVA and equal variances

Shih-Ni Prim

11/15/2020

We want to use a MANOVA test to see if the predictors `Time_Point`, `Drug`, and `Isotype` have any effects on the response variable vector that includes `H_CDR3`, `HMuFreq`, `L_CDR3`, `LMuFreq`, and `Binding`.

We first check assumptions. I beieve our large sample size n = 2464 allows us to bypass the normality assumption. Each row is assume to be independent. However, we run into some problems when we check the equal variance-covariance matrices assumption.

I ran the Fligner-Killeen Test of Homogeneity of Variances on all these variables. As shown in the output below, most of the p-values are very small, meaning the null hypothesis of equal variance is reject. In most cases, `H_CDR3`, `HMuFreq`, and `Binding` do not meet the equal variance-covariance matrices assumption.

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  Binding by Drug
## Fligner-Killeen:med chi-squared = 201.85, df = 2, p-value < 2.2e-16

##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  H_CDR3 by Drug
## Fligner-Killeen:med chi-squared = 13.463, df = 2, p-value = 0.001193

##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  HMuFreq by Drug
## Fligner-Killeen:med chi-squared = 48.86, df = 2, p-value = 2.456e-11

##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  L_CDR3 by Drug
## Fligner-Killeen:med chi-squared = 2.4113, df = 2, p-value = 0.2995

##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  LMuFreq by Drug
## Fligner-Killeen:med chi-squared = 0.17372, df = 2, p-value = 0.9168

##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  Binding by Isotype
## Fligner-Killeen:med chi-squared = 502.25, df = 4, p-value < 2.2e-16
```

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  H_CDR3 by Isotype
## Fligner-Killeen:med chi-squared = 53.604, df = 4, p-value = 6.37e-11

##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  HMuFreq by Isotype
## Fligner-Killeen:med chi-squared = 83.84, df = 4, p-value < 2.2e-16

##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  L_CDR3 by Isotype
## Fligner-Killeen:med chi-squared = 1.7399, df = 4, p-value = 0.7835

##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  LMuFreq by Isotype
## Fligner-Killeen:med chi-squared = 18.381, df = 4, p-value = 0.001039

##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  Binding by Time_Point
## Fligner-Killeen:med chi-squared = 687.02, df = 3, p-value < 2.2e-16

##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  H_CDR3 by Time_Point
## Fligner-Killeen:med chi-squared = 18.036, df = 3, p-value = 0.0004323

##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  HMuFreq by Time_Point
## Fligner-Killeen:med chi-squared = 50.662, df = 3, p-value = 5.775e-11

##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  L_CDR3 by Time_Point
## Fligner-Killeen:med chi-squared = 0.56825, df = 3, p-value = 0.9037

##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  LMuFreq by Time_Point
## Fligner-Killeen:med chi-squared = 16.585, df = 3, p-value = 0.0008601
```

I checked the variance-covariance matrix of the five response variables in different groups and look at the ratio (one matrix divided by another matrix). Some matrices seem quite different. For example, the ratio between variance-covariance matrices of Drug 2 and Drug 3 has numbers such as 8, 9, and 26.

```
##               H_CDR3      HMuFreq     L_CDR3    LMuFreq    Binding
## H_CDR3    0.82334669  0.09403869  1.7311285 -8.1369695  0.3921688
## HMuFreq   0.09403869  1.64598653 -1.2232736 26.2072877  0.8856344
## L_CDR3    1.73112854 -1.22327364  0.9656927  0.7827240 -9.1217203
## LMuFreq  -8.13696945 26.20728766  0.7827240  0.9973451  1.2938335
## Binding   0.39216879  0.88563437 -9.1217203  1.2938335  0.4772461
```

The ratio of variance-covariance matrices between Time_Point 0 and Time_Point 2 also has numbers greater than 4.

```
##               H_CDR3 HMuFreq     L_CDR3      LMuFreq      Binding
## H_CDR3    0.89777755      NA  4.1921601 -1.836937971 -0.047003143
## HMuFreq           NA      NA         NA           NA           NA
## L_CDR3    4.19216014      NA  1.0096036  0.342180121 -0.181909184
## LMuFreq  -1.83693797      NA  0.3421801  1.193364981  0.004007426
## Binding  -0.04700314      NA -0.1819092  0.004007426  0.003378924
```

The ratio of variance-covariance matrices between Isotype IgG and IgM even has a value of over 120.

```
##               H_CDR3 HMuFreq    L_CDR3      LMuFreq     Binding
## H_CDR3     1.6170960      NA 2.3899982   -0.6715703   16.432288
## HMuFreq           NA      NA        NA           NA          NA
## L_CDR3     2.3899982      NA 1.2285202    0.2902124    7.992429
## LMuFreq   -0.6715703      NA 0.2902124    0.6295988  120.242419
## Binding   16.4322878      NA 7.9924286  120.2424195   15.795280
```

Does this mean we can only use variables that have large p-values in the Fligner-Killeen Test of Homogeneity of Variance? It looks like maybe we can only use `L_CDR3` and `LMuFreq` in the response variable vector for the MANOVA test. Or is there a way to set `var.equal = FALSE` for MANOVA?