Multiple Comparison Procedures using R

AGH Practicals

P-value Adjustments

Garcia-Arenzana et al. (2014) tested associations of 25 dietary variables with mammographic density, an important risk factor for breast cancer, in Spanish women. They found the following results:

Dietary variable	**P-value**
Total calories	<0.001
Olive oil	0.008
Whole milk	0.039
White meat	0.041
Proteins	0.042
Nuts	0.060
Cereals and pasta	0.074
White fish	0.205
Butter	0.212
Vegetables	0.216
Skimmed milk	0.222
Red meat	0.251
Fruit	0.269
Eggs	0.275
Blue fish	0.340
Legumes	0.341
Carbohydrates	0.384
Potatoes	0.569
Bread	0.594
Fats	0.696
Sweets	0.762
Dairy products	0.940
Semi-skimmed milk	0.942
Total meat	0.975
Processed meat	0.986

Based on these results, five of the variables show a significant ($P < 0.05$) P value. However, Garcia-Arenzana et al. (2014) tested a total of 25 dietary variables. Hence, we'd expect one or two variables to show a significant result purely by chance, even if diet had no real effect on mammographic density. Using the Bonferroni correction, a test would have to have $P < 0.002$ to be significant. Under that criterion, only the test for total calories is significant. We will consider other error rates as well.

```
Input = ("
         Food               Raw.p
         Blue_fish         .34
         Bread             .594
         Butter            .212
         Carbohydrates     .384
         Cereals_and_pasta .074
         Dairy_products    .94
         Eggs              .275
         Fats              .696
         Fruit             .269
         Legumes           .341
         Nuts              .06
         Olive_oil         .008
         Potatoes          .569
         Processed_meat    .986
         Proteins           .042
         Red_meat           .251
         Semi-skimmed_milk  .942
         Skimmed_milk       .222
         Sweets             .762
         Total_calories     .001
         Total_meat         .975
         Vegetables         .216
         White_fish         .205
         White_meat         .041
         Whole_milk         .039
         ")
Data = read.table(textConnection(Input), header=TRUE)
```

Order data by p-value

```
Data = Data[order(Data$Raw.p),]
```

Perform p-value adjustments and add to data frame

```
Data$Bonferroni <- p.adjust(Data$Raw.p, method = "bonferroni")
Data$BH <- p.adjust(Data$Raw.p, method = "BH")
Data$Holm <- p.adjust(Data$ Raw.p, method = "holm")
Data$Hochberg <- p.adjust(Data$ Raw.p, method = "hochberg")
Data$BY <- p.adjust(Data$ Raw.p, method = "BY")
Data
```

```
##                 Food Raw.p Bonferroni        BH  Holm Hochberg         BY
## 20    Total_calories 0.001      0.025 0.0250000 0.025    0.025 0.09539895
## 12         Olive_oil 0.008      0.200 0.1000000 0.192    0.192 0.38159582
## 25        Whole_milk 0.039      0.975 0.2100000 0.897    0.882 0.80135122
## 24        White_meat 0.041      1.000 0.2100000 0.902    0.882 0.80135122
## 15          Proteins 0.042      1.000 0.2100000 0.902    0.882 0.80135122
```

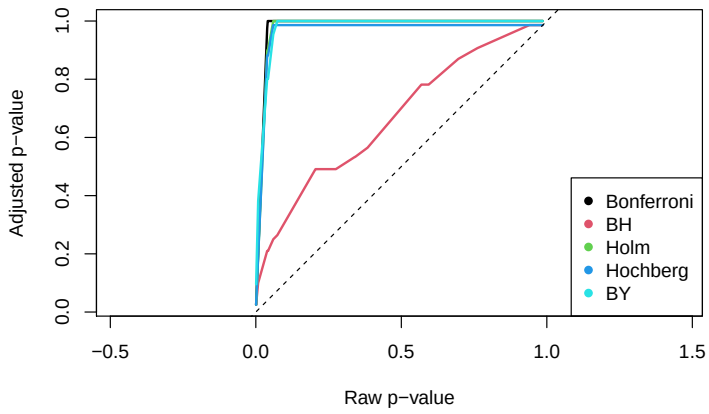
```
## 11              Nuts 0.060      1.000 0.2500000 1.000    0.986 0.95398954
## 5  Cereals_and_pasta 0.074      1.000 0.2642857 1.000    0.986 1.00000000
## 23        White_fish 0.205      1.000 0.4910714 1.000    0.986 1.00000000
## 3             Butter 0.212      1.000 0.4910714 1.000    0.986 1.00000000
## 22        Vegetables 0.216      1.000 0.4910714 1.000    0.986 1.00000000
## 18      Skimmed_milk 0.222      1.000 0.4910714 1.000    0.986 1.00000000
## 16          Red_meat 0.251      1.000 0.4910714 1.000    0.986 1.00000000
## 9              Fruit 0.269      1.000 0.4910714 1.000    0.986 1.00000000
## 7               Eggs 0.275      1.000 0.4910714 1.000    0.986 1.00000000
## 1          Blue_fish 0.340      1.000 0.5328125 1.000    0.986 1.00000000
## 10           Legumes 0.341      1.000 0.5328125 1.000    0.986 1.00000000
## 4      Carbohydrates 0.384      1.000 0.5647059 1.000    0.986 1.00000000
## 13          Potatoes 0.569      1.000 0.7815789 1.000    0.986 1.00000000
## 2              Bread 0.594      1.000 0.7815789 1.000    0.986 1.00000000
## 8               Fats 0.696      1.000 0.8700000 1.000    0.986 1.00000000
## 19            Sweets 0.762      1.000 0.9071429 1.000    0.986 1.00000000
## 6     Dairy_products 0.940      1.000 0.9860000 1.000    0.986 1.00000000
## 17 Semi-skimmed_milk 0.942      1.000 0.9860000 1.000    0.986 1.00000000
## 21        Total_meat 0.975      1.000 0.9860000 1.000    0.986 1.00000000
## 14    Processed_meat 0.986      1.000 0.9860000 1.000    0.986 1.00000000
```

Plot

```
X <- Data$Raw.p
Y <- cbind(Data$Bonferroni, Data$BH, Data$Holm, Data$Hochberg, Data$BY)

matplot(X, Y, xlab="Raw p-value", ylab="Adjusted p-value",
        type="l", asp=1, col=1:6, lty=1, lwd=2)

legend('bottomright', legend = c("Bonferroni", "BH", "Holm", "Hochberg", "BY"),
       col = 1:5, cex = 1, pch = 16)
abline(0, 1, col=1, lty=2, lwd=1)
```



Hypothetical Example

```
Input = ("
Factor   Raw.p
 A        .001
 B        .01
 C        .025
 D        .05
 E        .1
")

Data = read.table(textConnection(Input),header=TRUE)
```

Perform p-value adjustments and add to data frame

```
Data$Bonferroni <- p.adjust(Data$Raw.p, method = "bonferroni")
Data$BH <- signif(p.adjust(Data$Raw.p, method = "BH"), 4)
Data$Holm <- p.adjust(Data$ Raw.p, method = "holm")
Data$Hochberg <- p.adjust(Data$ Raw.p, method = "hochberg")
Data$BY <- signif(p.adjust(Data$ Raw.p, method = "BY"), 4)
Data
```

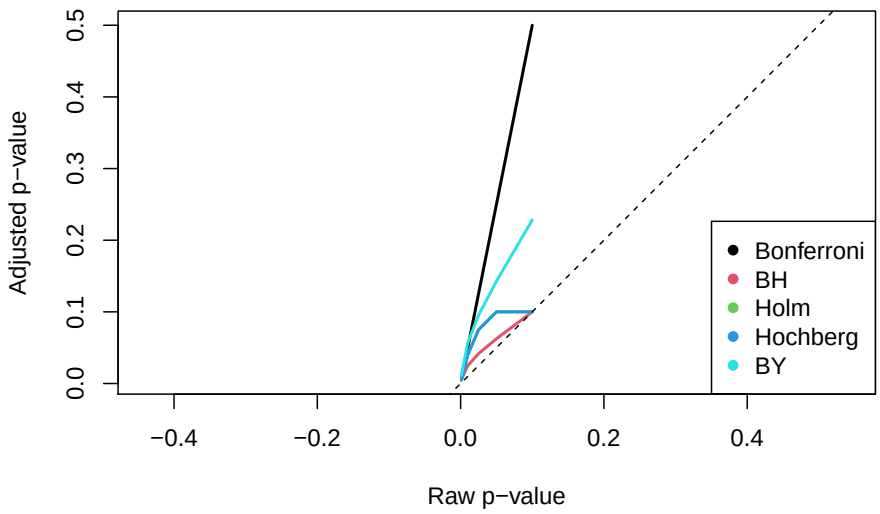
```
##   Factor Raw.p Bonferroni      BH  Holm Hochberg      BY
## 1      A 0.001      0.005 0.00500 0.005    0.005 0.01142
## 2      B 0.010      0.050 0.02500 0.040    0.040 0.05708
## 3      C 0.025      0.125 0.04167 0.075    0.075 0.09514
## 4      D 0.050      0.250 0.06250 0.100    0.100 0.14270
## 5      E 0.100      0.500 0.10000 0.100    0.100 0.22830
```

Plot

```
X <- Data$Raw.p
Y <- cbind(Data$Bonferroni, Data$BH, Data$Holm, Data$Hochberg, Data$BY)

matplot(X, Y, xlab="Raw p-value", ylab="Adjusted p-value", type="l", asp=1,
        col=1:6, lty=1, lwd=2)

legend('bottomright', legend = c("Bonferroni", "BH", "Holm", "Hochberg", "BY"),
       col = 1:5, cex = 1, pch = 16)
abline(0, 1, col=1, lty=2, lwd=1)
```



Source Code

```
p.adjust.methods <- c("holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none")

p.adjust <- function(p, method = p.adjust.methods, n = length(p))
{
    method <- match.arg(method)
    if(method == "fdr") method <- "BH"
    nm <- names(p)
    p <- as.numeric(p)
    p0 <- setNames(p, nm)
    if(all(nna <- !is.na(p))) nna <- TRUE
    p <- p[nna]
    lp <- length(p)
    stopifnot(n >= lp)
    if (n <= 1) return(p0)
    if (n == 2 && method == "hommel") method <- "hochberg"

    p0[nna] <- switch(method, bonferroni = pmin(1, n * p),
         holm = {
            i <- seq_len(lp)
            o <- order(p)
            ro <- order(o)
            pmin(1, cummax( (n - i + 1L) * p[o] ))[ro]
         },
         hommel = { ## needs n-1 >= 2 in for() below
           if(n > lp) p <- c(p, rep.int(1, n-lp))
            i <- seq_len(n)
            o <- order(p)
```

```
            p <- p[o]
            ro <- order(o)
            q <- pa <- rep.int( min(n*p/i), n)
           for (j in (n-1):2) {
            ij <- seq_len(n-j+1)
            i2 <- (n-j+2):n
            q1 <- min(j*p[i2]/(2:j))
            q[ij] <- pmin(j*p[ij], q1)
            q[i2] <- q[n-j+1]
            pa <- pmax(pa,q)
           }
           pmax(pa,p)[if(lp < n) ro[1:lp] else ro]
           },
         hochberg = {
            i <- lp:1L
            o <- order(p, decreasing = TRUE)
            ro <- order(o)
            pmin(1, cummin( (n - i + 1L) * p[o] ))[ro]
         },
         BH = {
            i <- lp:1L
            o <- order(p, decreasing = TRUE)
            ro <- order(o)
            pmin(1, cummin( n / i * p[o] ))[ro]
         },
         BY = {
            i <- lp:1L
            o <- order(p, decreasing = TRUE)
            ro <- order(o)
            q <- sum(1L/(1L:n))
            pmin(1, cummin(q * n / i * p[o]))[ro]
         },
          none = p)
    p0
}

bonf <- p.adjust(Data$Raw.p, method = "bonferroni")
bonf
```

```
## [1] 0.005 0.050 0.125 0.250 0.500
```

References

Garcia-Arenzana, N., E.M. Navarrete-Munoz, V. Lope, P. Moreo, S. Laso-Pablos, N. Ascunce, F. Casanova-Gomez, C. Sanchez-Contador, C. Santamaria, N. Aragones, B.P. Gomez, J. Vioque, and M. Pollon. 2014. Calorie intake, olive oil consumption and mammographic density among Spanish women. International Journal of Cancer 134: 1916-1925.

McDonald, J.H. 2014. Handbook of Biological Statistics (3rd ed.). Sparky House Publishing, Baltimore, Maryland