PS3

Shuai Yan

2023-10-25

To create reproducible results, set random seed to 1.

```
set.seed(19710822)
```

Part I

```
myoglobin <- data.frame(s = c(1.1,  1.5,  1.6,  2.3,  3.4,  5.3,  7.5,  8.4, 14.1),
                        v = c(1.49, 1.79, 1.79, 2.11, 2.83, 3.42, 3.79, 3.97, 4.08))

# models
michaelis_menten_model <- nls(v ~ Vmax * s / (Km + s),
                          start = list(Vmax = 4, Km = 2),
                          data = myoglobin)

non_specific_binding_model <- nls(v ~ (Vmax * s / (Km + s)) + Mns * s,
                              start = list(Vmax = 4, Km = 2, Mns = 1),
                              data = myoglobin)

background_signal_model <- nls(v ~ (Vmax * s / (Km + s)) + Vbk,
                              start = list(Vmax = 4, Km = 2, Vbk = 1),
                              data = myoglobin)

ns_bk_model <- nls(v ~ (Vmax * s / (Km + s)) + Mns*s + Vbk,
                    start = list(Vmax = 4, Km = 2, Mns = 0.5, Vbk = 0.5),
                    data = myoglobin)

# AIC correction
corr_AIC <- function(AIC_value, n, P) {
  return(AIC_value + 2 * (P + 1) * (P + 2) / (n - P))
}

# corrected AIC
n <- length(myoglobin$s)
P_mm <- length(coefficients(michaelis_menten_model))
P_ns <- length(coefficients(non_specific_binding_model))
P_bk <- length(coefficients(background_signal_model))
P_ns_bk <- length(coefficients(ns_bk_model))

AIC_mm <- corr_AIC(AIC(michaelis_menten_model), n, P_mm)
AIC_ns <- corr_AIC(AIC(non_specific_binding_model), n, P_ns)
AIC_bk <- corr_AIC(AIC(background_signal_model), n, P_bk)
AIC_ns_bk <- corr_AIC(AIC(ns_bk_model), n, P_ns_bk)
```

```
delta_ns <- AIC_ns - AIC_mm
delta_bk <- AIC_bk - AIC_mm
delta_ns_bk <- AIC_ns_bk - AIC_mm
```

Compared to the Michaelis Menten model, ΔAIC of the non-specific binding model is: -0.7753711; ΔAIC of the background signal model is: 4.9868849; ΔAIC of the model with both terms is: -0.5574722. None of them is less then -6, thus there is not enough evidence to choose either of the more complex models.

Part II

```
control <- data.frame(    t = rnorm(10, mean = 7,  sd = 0.6),
                      group = factor("ctrl"))
drug_a  <- data.frame(    t = rnorm( 8, mean = 9,  sd = 0.6),
                      group = factor("drgA"))
drug_b  <- data.frame(    t = rnorm( 9, mean = 7,  sd = 0.6),
                      group = factor("drgB"))
drug_c  <- data.frame(    t = rnorm( 7, mean = 7,  sd = 0.6),
                      group = factor("drgC"))
drug_d  <- data.frame(    t = rnorm( 8, mean = 11, sd = 0.6),
                      group = factor("drgD"))
d <- rbind(control, drug_a, drug_b, drug_c, drug_d)

a_test <- t.test(x=drug_a$t, y=control$t, conf.level=0.05)
b_test <- t.test(x=drug_b$t, y=control$t, conf.level=0.05)
c_test <- t.test(x=drug_c$t, y=control$t, conf.level=0.05)
d_test <- t.test(x=drug_d$t, y=control$t, conf.level=0.05)
```

```
print(a_test)
```

```
##
##  Welch Two Sample t-test
##
## data:  drug_a$t and control$t
## t = 4.7674, df = 14.103, p-value = 0.0002945
## alternative hypothesis: true difference in means is not equal to 0
## 5 percent confidence interval:
##  1.544872 1.586802
## sample estimates:
## mean of x mean of y
##  8.902544  7.336707
```

```
print(b_test)
```

```
##
##  Welch Two Sample t-test
##
## data:  drug_b$t and control$t
## t = -2.7152, df = 16.318, p-value = 0.01509
## alternative hypothesis: true difference in means is not equal to 0
## 5 percent confidence interval:
##  -0.7085626 -0.6760893
## sample estimates:
## mean of x mean of y
##  6.644381  7.336707
```

```r
print(c_test)
```

```
##
##  Welch Two Sample t-test
##
## data:  drug_c$t and control$t
## t = -1.8301, df = 13.771, p-value = 0.08895
## alternative hypothesis: true difference in means is not equal to 0
## 5 percent confidence interval:
##  -0.5697621 -0.5313411
## sample estimates:
## mean of x mean of y
##  6.786155  7.336707
```

```r
print(d_test)
```

```
##
##  Welch Two Sample t-test
##
## data:  drug_d$t and control$t
## t = 14.674, df = 15.681, p-value = 1.416e-10
## alternative hypothesis: true difference in means is not equal to 0
## 5 percent confidence interval:
##  3.720163 3.752612
## sample estimates:
## mean of x mean of y
## 11.073094  7.336707
```

Thus drug A, drug B, and drug D are significant.

```r
p_raw <- c(a_test$p.value,
           b_test$p.value,
           c_test$p.value,
           d_test$p.value)
print(which(p.adjust(p_raw, method="BH") < 0.05))
```

```
## [1] 1 2 4
```

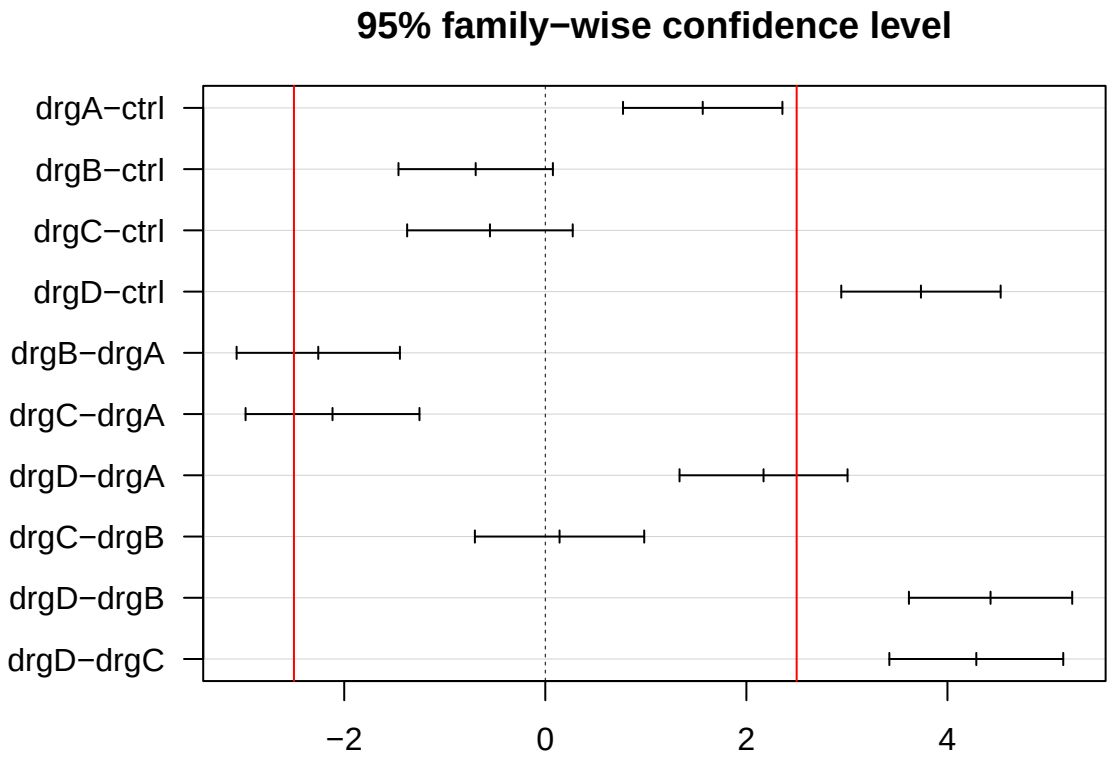
Thus, our result is not changed after Benjamini Hochberg correction.

```r
anova_result <- aov(t ~ group, data=d)
tHSD <- TukeyHSD(anova_result)
print(tHSD)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = t ~ group, data = d)
##
## $group
##                 diff        lwr        upr     p adj
## drgA-ctrl  1.5658368  0.7732179  2.3584557 0.0000171
## drgB-ctrl -0.6923259 -1.4600917  0.0754398 0.0940236
## drgC-ctrl -0.5505516 -1.3740232  0.2729200 0.3266882
## drgD-ctrl  3.7363877  2.9437687  4.5290066 0.0000000
## drgB-drgA -2.2581627 -3.0701167 -1.4462088 0.0000000
## drgC-drgA -2.1163884 -2.9812066 -1.2515703 0.0000003
```

```
## drgD-drgA  2.1705509  1.3350571  3.0060446 0.0000001
## drgC-drgB  0.1417743 -0.7003242  0.9838728 0.9884934
## drgD-drgB  4.4287136  3.6167597  5.2406676 0.0000000
## drgD-drgC  4.2869393  3.4221211  5.1517575 0.0000000
```

```
# plot
par(mar=c(5,8,2,1)) ; plot(tHSD, las=1) # horizontal x-axis labels
abline(v = c(-2.5, 2.5), col="red")
```



Only drug A and drug D are significant based on ANOVA and Tukey's Honest Significant Differences post-testing procedure.

Based on our calculation, I prefer omnibus test followed by post-testing procedure. The reason is that this test provides more power over multiple T tests, which is mostly persued except for screening. Only if the probem I am facing is clearly a screening problem, in which I want more positive data regardless of more false positive, I may use multiple T tests instead. In addition, there is a higher risk of forgetting correction after multiple T tests, which is not a problem with my preference.