5. ANOVA (Analysis of Variance)

[ANOVA 1] One-way ANOVA ⇒ Fixed effect regression model

(1) $Y_{ij} = \mu_i + \epsilon_{ij}$ $\quad i = 1 \cdots K, \; j = 1, \cdots, n_i$

(2) $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ $\quad \sum_{i=1}^{K} \alpha_i = 0$

[ANOVA 2] (2.1) One-way ANOVA global F-test

Under $H_0 : \mu_1 = \mu_2 = \cdots = \mu_K \iff H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_K$

$$F_{stat} = \frac{(RSS_H - RSS_{Full}) / (K-1)}{RSS_{Full} / (n-K)} \sim F_{K-1, n-K}$$

$$RSS_{Full} = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\cdot})^2 \qquad \hat{\mu}_i = \bar{Y}_{i\cdot}$$

$$RSS_H = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{\cdot\cdot})^2 \qquad \hat{\mu}_H = \bar{Y}_{\cdot\cdot}$$

(2.2) ANOVA table

Source	Sum of Squares	Degrees of freedom
Between groups	$SS_{Between} = \sum_{i=1}^{K} n_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	$K-1$
Within groups	$SS_{Within} = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\cdot})^2$	$n-K$
Total	$SS_{Total} = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{\cdot\cdot})^2$	$n-1$

△ When the global F-test is reject ⇒ look for contrasts

[ANOVA 3] One contrast

① Hypothesis test : under $H_0 : \mu_i = \mu_j$

$$\frac{\hat{\mu}_i - \hat{\mu}_j}{\text{S.E.}(\hat{\mu}_i - \hat{\mu}_j)} \sim t_{df} \quad \text{with} \quad \text{S.E.}(\hat{\mu}_i - \hat{\mu}_j) = \hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

$$df = n - k$$

② Confidence interval for $\mu_i - \mu_j$

$$\hat{\mu}_i - \hat{\mu}_j \pm t_{df}^{(\alpha/2)} \times \text{S.E.}(\hat{\mu}_i - \hat{\mu}_j)$$

△ K groups ⇒ $\binom{k}{2}$ pairwise comparisons ⇒ multiple contrasts

[ANOVA 4] Multiple contrasts (Simultaneous inference)

Given multiple hypotheses $H_{01}, \cdots, H_{0m}$ for parameters $\theta_1 \cdots \theta_m$

(4.1) Familywise Error Rate (FWER) is probability of rejecting at least one of $H_{01}, \cdots, H_{0m}$ when they are all true

(4.2) Simultaneous confidence intervals at $100(1-\alpha)\%$ level are intervals (L_i, U_i) $i = 1 \cdots m$ with

$$P[\, L_i \le \theta_i \le U_i \text{ for all } i = 1, \cdots, m \,] > 1 - \alpha$$

(4.3) Bonferroni's correction Let p_i be p-value of H_{0i}.

① Hypothesis test :

V1 : reject H_{0i} if $p_i < \frac{\alpha}{m}$

V2: define $p_{i,adj} = \min\{m \times p_i, 1\}$, reject H_{0i} if $p_{i,adj} < \alpha$

② Simultaneous confidence intervals: Change quantile to at level $\frac{\alpha}{2m}$

E.g. $\hat{\mu}_i - \hat{\mu}_j \pm t_{df}^{(\frac{\alpha}{2m})} \times S.E.(\hat{\mu}_i - \hat{\mu}_j)$ with $m = \binom{K}{2}$, $df = n-K$

Differences and connections between ① one-way ANOVA F-test ② one contrast t-test ③ multiple contrast tests

(4.4) Tukey-Kramer procedure (for pairwise comparisions)

Use under balanced design and $H_{0,ij}: \mu_i = \mu_j$ all hold,

$$\max_{1 \le i < j \le K} |t_{ij}| = \max_{1 \le i < j \le K} \frac{|\bar{y}_{i\cdot} - \bar{y}_{j\cdot}|}{\hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \sim \frac{1}{\sqrt{2}} q_{K, n-K}$$

studentized distribution with $(K, n-K)$ parameters

① Hypothesis test: reject $H_{0,ij}: \mu_i = \mu_j$ ✓

if $|t_{ij}| > \frac{1}{\sqrt{2}} q_{K,n-K}^{(\alpha)}$ ✓

② Tukey's HSD intervals ✓

$$\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \pm \frac{1}{\sqrt{2}} q_{K,n-K}^{(\alpha)} \times \hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

(4.5) False Discovery Rate (FDR)

$$FDR = E(FDP) = E\left\{\frac{\#\text{ False Discoveries}}{\#\text{ Discoveries}}\right\}$$ ✓

Benjamini - Hochberg (BH) procedure

controls FDR $\leq \frac{m_0}{m} \alpha \leq \alpha$

for independent tests

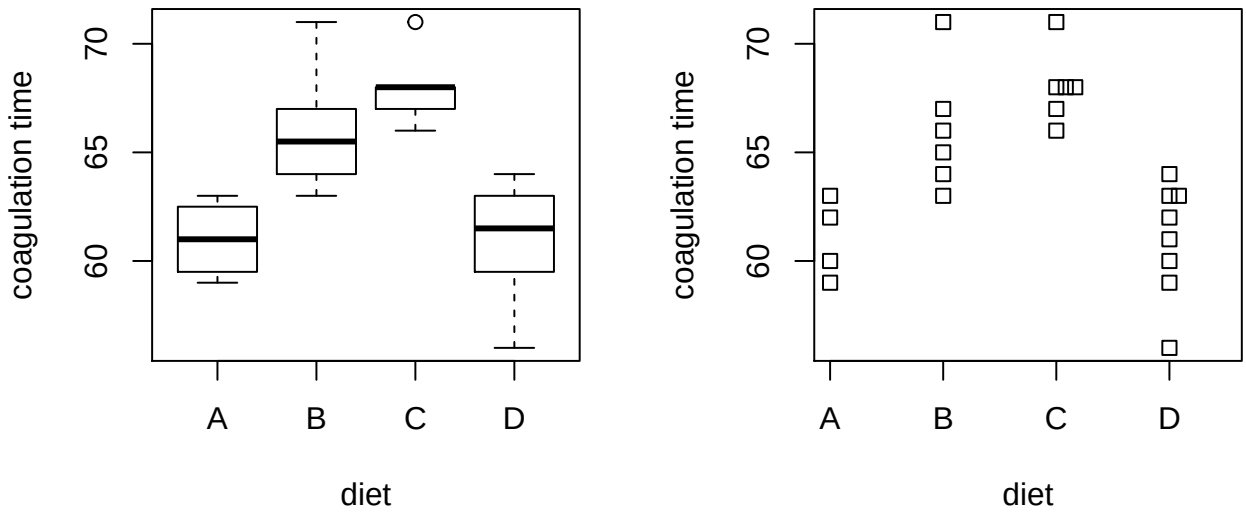
Example on One-Way ANOVA

- 24 animals were randomly assigned to four different diets and
- The blood coagulation time was measured. Box et al. (1978).

```
library(faraway)
data(coagulation, package="faraway")
head(coagulation)
```

```
##   coag diet
## 1   62    A
## 2   60    A
## 3   63    A
## 4   59    A
## 5   63    B
## 6   67    B
```

```
par(mfrow=c(1,2))
plot(coag ~ diet, coagulation,ylab="coagulation time")
stripchart(coag ~ diet, coagulation, vertical=TRUE, method="stack",
           xlab="diet",ylab="coagulation time")
```



```
par(mfrow=c(1,1))
```

- Left: boxplot.
- Right: stripchart. (1-dim scatterplot, an alternative to boxplots when sample sizes are small.)
- Median and upper quartile of diet C are the same.
- There are ties in diets C and D.

1.ANOVA Global F-Test: $\mu_A = \mu_B = \mu_C = \mu_D$

ANOVA code version 1

```
lmodi <- lm(coag ~ diet -1, coagulation)
summary(lmodi)$coefficients
```

```
##       Estimate Std. Error  t value     Pr(>|t|)
## dietA       61  1.1832160 51.55441 9.547815e-23
## dietB       66  0.9660918 68.31649 3.532325e-25
## dietC       68  0.9660918 70.38669 1.948886e-25
## dietD       61  0.8366600 72.90895 9.663048e-26
```

```
lmnull <- lm(coag ~ 1, coagulation)
anova(lmnull,lmodi)
```

```
## Analysis of Variance Table
##
## Model 1: coag ~ 1
## Model 2: coag ~ diet - 1
##   Res.Df RSS Df Sum of Sq      F    Pr(>F)
## 1     23 340
## 2     20 112  3       228 13.571 4.658e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- We see that there is indeed a difference in the levels.

ANOVA code version 2

```
lmod <- lm(coag ~ diet, coagulation)
summary(lmod)$coefficients
```

```
##                 Estimate Std. Error      t value     Pr(>|t|)
## (Intercept) 6.100000e+01   1.183216 5.155441e+01 9.547815e-23
## dietB       5.000000e+00   1.527525 3.273268e+00 3.802505e-03
## dietC       7.000000e+00   1.527525 4.582576e+00 1.805132e-04
## dietD       2.991428e-15   1.449138 2.064281e-15 1.000000e+00
```

```
anova(lmod)
```

```
## Analysis of Variance Table
##
## Response: coag
##           Df Sum Sq Mean Sq F value    Pr(>F)
## diet       3    228    76.0  13.571 4.658e-05 ***
## Residuals 20    112     5.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note

```
anova(lmnull, lmod) #This is also ok.
```

```
## Analysis of Variance Table
##
## Model 1: coag ~ 1
## Model 2: coag ~ diet
##   Res.Df RSS Df Sum of Sq      F    Pr(>F)
## 1     23 340
## 2     20 112  3       228 13.571 4.658e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lmodi) #This is incorrect
```

```
## Analysis of Variance Table
##
## Response: coag
##           Df Sum Sq Mean Sq F value    Pr(>F)
## diet       4  98532 24633.0  4398.8 < 2.2e-16 ***
## Residuals 20    112     5.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Global test answers whether there is a significant difference between any of the levels.
- This test does not tell us which levels are different from others.

2. ANOVA: Pairwise Comparisons

- A pairwise comparison test $H_0 : \mu_i = \mu_j$ for two levels i and j in $\{1, \ldots, K\}$.
- This can be done by a t-test as under $H_0 : \mu_i = \mu_j$:

$$\frac{\hat{\mu}_i - \hat{\mu}_j}{\text{s. e.}\,(\hat{\mu}_i - \hat{\mu}_j)} \sim t_{df}, \qquad \text{s. e.}\,(\hat{\mu}_i - \hat{\mu}_j) = \hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}, \qquad df = n - K.$$

- Or can be made using a CI for $\mu_i - \mu_j$ using:

$$\hat{\mu}_i - \hat{\mu}_j \pm t_{df}^{(\alpha/2)} \times \text{s. e.}\,(\hat{\mu}_i - \hat{\mu}_j)\,, \qquad \text{s. e.}\,(\hat{\mu}_i - \hat{\mu}_j) = \hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}, \quad df = n - K.$$

  where $H_0 : \mu_i = \mu_j$ is rejected at the significance level α if 0 is not contained in the interval.

- Example: Consider the `coagulation` data. Suppose we test $\mu_A = \mu_B$

```
lmodi <- lm(coag ~ diet -1, coagulation)
(tmp_table <- summary(lmodi)$coefficients)
```

```
##       Estimate Std. Error  t value     Pr(>|t|)
## dietA       61  1.1832160 51.55441 9.547815e-23
## dietB       66  0.9660918 68.31649 3.532325e-25
## dietC       68  0.9660918 70.38669 1.948886e-25
## dietD       61  0.8366600 72.90895 9.663048e-26
```

```
(muB_muA <- tmp_table["dietB","Estimate"] - tmp_table["dietA","Estimate"] )
```

```
## [1] 5
```

```
(nA <- sum(coagulation$diet == "A")) #A level sample size
```

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

```
(nB <- sum(coagulation$diet == "B")) #B level sample size
```

```
## [1] 6
```

```
(s.e.muB_muA <-  summary(lmodi)$sigma * sqrt(1/nA + 1/nB)) #standard error
```

```
## [1] 1.527525
```

```r
muB_muA /s.e.muB_muA # t statistic
```

```
## [1] 3.273268
```

```r
2 * pt( abs(muB_muA /s.e.muB_muA ), df = 24-4 , lower.tail=FALSE )
```

```
## [1] 0.003802505
```

```r
#A simplier way to see all pairwise comparison
pairwise.t.test(coagulation$coag, coagulation$diet, p.adj = "none" )
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  coagulation$coag and coagulation$diet
##
##   A       B       C
## B 0.00380 -       -
## C 0.00018 0.15878 -
## D 1.00000 0.00086 2.3e-05
##
## P value adjustment method: none
```

3. Multiplicity Adjustment for all pairwise comparisons

3.1 Bonferroni's correction

```
pairwise.t.test(coagulation$coag, coagulation$diet, p.adj = "bonferroni" )
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  coagulation$coag and coagulation$diet
##
##   A       B       C
## B 0.02282 -       -
## C 0.00108 0.95266 -
## D 1.00000 0.00518 0.00014
##
## P value adjustment method: bonferroni
```

A simulaiton comparison between F-test and Bonferroni's correction

- Step 1: write a function to generate data

```
#generate data with K levels factor
K = 5
all_mu_values <- rep(1,K) #global null, all mean are equal to 1
set.seed(123)
#generate data with specified mean values
generate_balanced_data <- function(K, all_mu_values){
  all_sample_sizes <- rep(20, K) #sample sizes of each level
  n_total = sum(all_sample_sizes)
  factor_var <- rep(LETTERS[1:K], all_sample_sizes)
  y_all = NULL
  for(j in 1:K){
    yj <- rnorm( n = all_sample_sizes[j], mean = all_mu_values[j], sd = 1)
    y_all <- c(y_all, yj)
  }
```

```
  my_data <- data.frame( y_all, factor_var)
  colnames(my_data) <- c("y", "factor")
  return(my_data)
}
my_data <- generate_balanced_data(K, all_mu_values)
head(my_data, 3)
```

```
##           y factor
## 1 0.4395244      A
## 2 0.7698225      A
## 3 2.5587083      A
```

- Step 2: write a function that conducts F-test and Bonferroni's procedure

```
alpha_level = 0.05
tests_ANOVA_F_and_bonf <- function(my_data,alpha_level){
  #1. one-way ANOVA F-test
  lmod <- lm(y ~ factor, my_data)
  p_val_F_test <- anova(lmod)$`Pr(>F)`[1] #read p-value of one-way ANOVA
  rej_F_test <- (p_val_F_test < alpha_level)

  #2. all pairwise t-test with Bonferrnoi correction
  pair_bonf_res <- pairwise.t.test(my_data$y, my_data$factor,
                                   p.adj = "bonferroni" )$p.value
  rej_pair_bonf <-  any(na.omit(pair_bonf_res) < alpha_level)
  #na.omit drop NA to see if any rejection

  return(c(rej_F_test, rej_pair_bonf) )
}
tests_ANOVA_F_and_bonf(my_data,alpha_level)
```

```
## [1] FALSE FALSE
```

- Step 3: conduct repeated simulations to estimate type I error under $H_0 : \mu_A = \mu_B = \mu_C = \mu_D = \mu_E$.

```r
set.seed(123)
N_repeat_number <- 1000
all_rej_F_test = rep(NA, N_repeat_number)
all_rej_t_bonf = rep(NA, N_repeat_number)
for(rep_ind in 1:N_repeat_number){
  my_data <- generate_balanced_data(K, all_mu_values)
  test_res <- tests_ANOVA_F_and_bonf(my_data,alpha_level)
  all_rej_F_test[rep_ind] <- test_res[1]
  all_rej_t_bonf[rep_ind] <- test_res[2]
}
#Estimated Type I error of ANOVA F-test
mean(all_rej_F_test)
```

```
## [1] 0.052
```

```r
#Estimated Type I error of all pairwise t-tests with Bonferrnoi correction
mean(all_rej_t_bonf) #smaller than nominal level 0.05
```

```
## [1] 0.023
```

3.2 Tukey's honest significant difference (HSD)

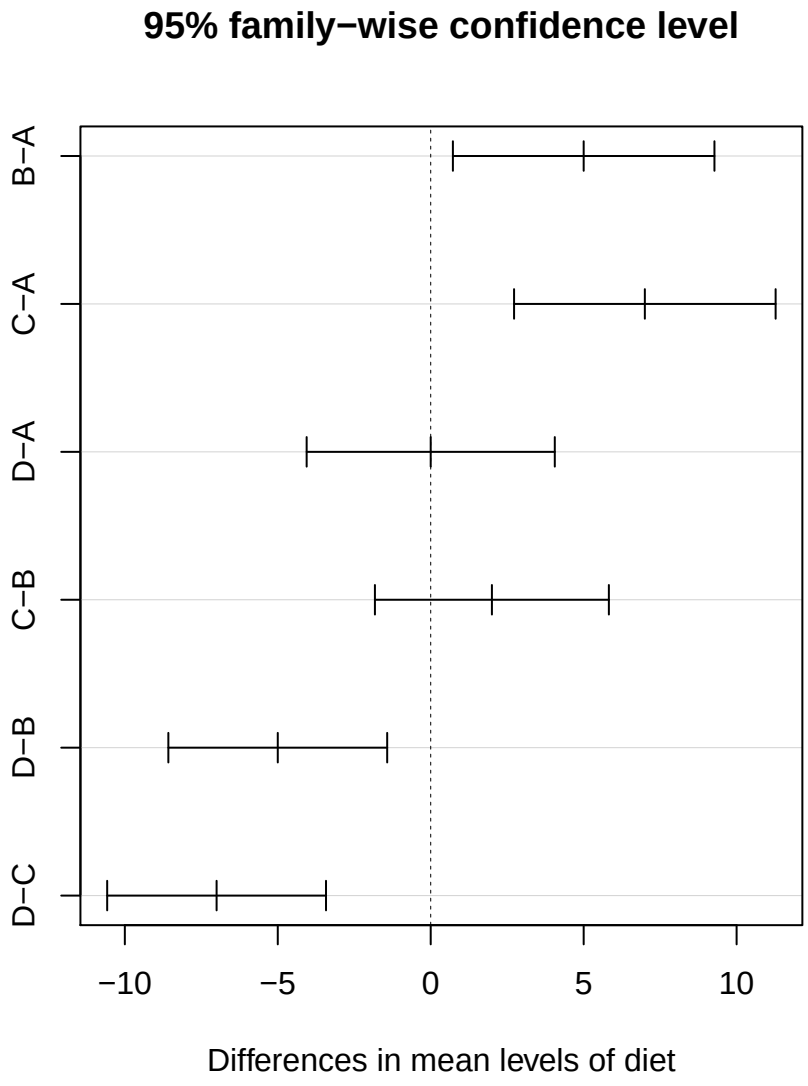
- The Tukey CIs are:

$$\hat{\mu}_i - \hat{\mu}_j \pm \frac{q_\alpha(K, df)}{\sqrt{2}} \times \hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

```
#A convenient way to obtain all the intervals is:
anova_res <- aov(coag ~ diet, coagulation)
(tci <- TukeyHSD(anova_res))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = coag ~ diet, data = coagulation)
##
## $diet
##     diff         lwr       upr     p adj
## B-A    5   0.7245544  9.275446 0.0183283
## C-A    7   2.7245544 11.275446 0.0009577
## D-A    0  -4.0560438  4.056044 1.0000000
## C-B    2  -1.8240748  5.824075 0.4766005
## D-B   -5  -8.5770944 -1.422906 0.0044114
## D-C   -7 -10.5770944 -3.422906 0.0001268
```

```
plot(tci)
```



A simulaiton on Tukey's method under balanced and unbalanced designs

- Step 1: data generate functions

```r
#generate data with K levels factor
K = 5
all_mu_values <- rep(1,K) #global null, all mean are equal to 1
set.seed(123)
#Balanced design: generate data with specified mean values
generate_balanced_data <- function(K, all_mu_values){
  all_sample_sizes <- rep(20, K) #sample sizes of each level
  n_total = sum(all_sample_sizes)
  factor_var <- rep(LETTERS[1:K], all_sample_sizes)
```

```
  y_all = NULL
  for(j in 1:K){
    yj <- rnorm( n = all_sample_sizes[j], mean = all_mu_values[j], sd = 1)
    y_all <- c(y_all, yj)
  }
  my_data <- data.frame( y_all, factor_var)
  colnames(my_data) <- c("y", "factor")
  return(my_data)
}
#Unbalanced design: generate data with specified mean values
generate_unbalanced_data <- function(K, all_mu_values){
  all_sample_sizes <- c(5,5,5, 30, 50)
  n_total = sum(all_sample_sizes)
  factor_var <- rep(LETTERS[1:K], all_sample_sizes)
  y_all = NULL
  for(j in 1:K){
    yj <- rnorm( n = all_sample_sizes[j], mean = all_mu_values[j], sd = 1)
    y_all <- c(y_all, yj)
  }
  my_data <- data.frame( y_all, factor_var)
  colnames(my_data) <- c("y", "factor")
  return(my_data)
}
```

- Step 2: write a function that conducts test by Tukey's adjustment method

```
alpha_level = 0.05
tests_ANOVA_tukey <- function(my_data,alpha_level){
  tci <- TukeyHSD( aov(y ~ factor , my_data))
  rej_pair_Tukey <- any(tci$factor[,"p adj"] < alpha_level)
  return(rej_pair_Tukey)
}
```

- Step 3: conduct repeated simulations to estimate type I error under $H_0 : \mu_A = \mu_B =$

$\mu_C = \mu_D = \mu_E$.

```
set.seed(123)
N_repeat_number <- 1000
all_rej_tukey_balanced = rep(NA, N_repeat_number)
all_rej_tukey_unbalanced = rep(NA, N_repeat_number)
for(rep_ind in 1:N_repeat_number){
  #balanced design
  my_data <- generate_balanced_data(K, all_mu_values)
  test_res1 <- tests_ANOVA_tukey(my_data,alpha_level)
  all_rej_tukey_balanced[rep_ind] <- test_res1

  #unbalanced design
  my_data <- generate_unbalanced_data(K, all_mu_values)
  test_res2 <- tests_ANOVA_tukey(my_data,alpha_level)
  all_rej_tukey_unbalanced[rep_ind] <- test_res2
}
#Estimated Type I error of tukey's method for balanced design
mean(all_rej_tukey_balanced)
```

```
## [1] 0.052
```

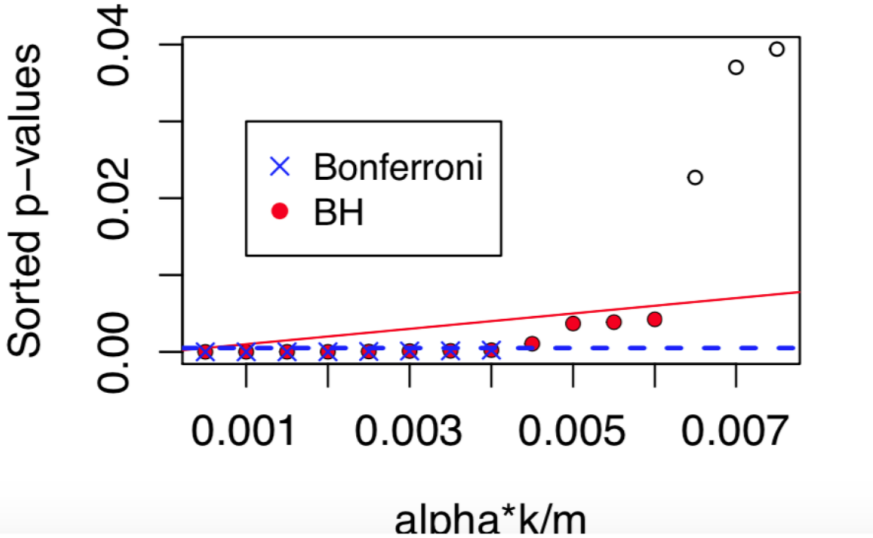
```
#Estimated Type I error of tukey's method for unbalanced design
mean(all_rej_tukey_unbalanced)
```

```
## [1] 0.035
```

```
#Estimated type I error under the balanced design is closer to 0.05
```

3.3 Pairwise comparison with BH procedure

- An illustraion of compmarison between BH and bonferroni's correction




```
pairwise.t.test(coagulation$coag, coagulation$diet, p.adj = "BH" )
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  coagulation$coag and coagulation$diet
##
##   A       B       C
## B 0.00570 -       -
## C 0.00054 0.19053 -
## D 1.00000 0.00173 0.00014
##
## P value adjustment method: BH
```

- Tukey's CI is specific to all pairwise comparisons.
- But the other multiplicity adjustment methods (Bonferroni's correctiona and BH) are not restricted to pairwise comparisons and can also be used in general settings.
- Given a general set of p-values, the function `p.adjust()` can be used to obtain adjusted p-values by setting the option `method = "bonferroni"` or `method = "BH"`.