# Multiple Comparisons

If the F test in One-Way ANOVA shows the population means are different, then often we want to further examine which means differ. A common way is to make pairwise comparisons.
If there are $t$ treatments, there are $\binom{t}{2} = \frac{t!}{2!(t-2)!}$ pairwise comparisons.
e.g, if t=3, there are 3 pairwise comparisons (1 vs 2, 1vs 3 and 2 vs 3) and if t=4, there are 6 pairwise comparisons (1 vs 2, 1 vs 3, 1 vs 4, 2 vs 3, 2 vs 4 and 3 vs 4).
Compare group $i$ to $j$:
$H_0 : \mu_i = \mu_j, H_a : \mu_i \neq \mu_j$.
$t = \frac{\bar{y}_{i\cdot} - \bar{y}_{j\cdot}}{\sqrt{MSE}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}$.
$t \sim t_\nu$ with $\nu = N - t$ under $H_0$.
$100(1 - \alpha)\%$ CI for $\mu_i - \mu_j$ is
$\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \pm t_{\frac{\alpha}{2};\nu}\sqrt{MSE}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$.
Note MSE is based on t samples with d.f.=N-t.

# One factor completely randomized design

example 5.1
Three covers on the box of cereal, 18 markets selected.

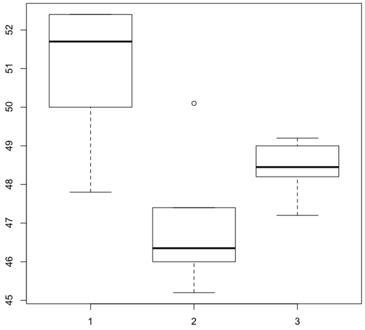
```
Cover 1: Sports hero: 52.4, 47.8, 52.4, 51.3, 50.0, 52.1
Cover 2: Child:       50.1, 45.2, 46.0, 46.5, 47.4, 46.2
Cover 3: Cereal Bowl: 49.2, 48.3, 49.0, 47.2, 48.6, 48.2
```

Is there a difference among the population means? Use $\alpha = 0.05$.
If there is a significant difference, get 95% CI for
$\mu_1 - \mu_2, \mu_1 - \mu_3, \mu_2 - \mu_3$ to see which means differ.

# boxplot of data

```
> cover1=c(52.4, 47.8, 52.4, 51.3, 50.0, 52.1)
> cover2=c(50.1, 45.2, 46.0, 46.5, 47.4, 46.2)
> cover3=c( 49.2, 48.3, 49.0, 47.2, 48.6, 48.2)

> boxplot(cover1,cover2,cover3)
```



```
> y=c(cover1,cover2,cover3)
> treatment=c(rep(1,6),rep(2,6),rep(3,6))
> output=aov(y~factor(treatment))
> summary(output)
                  Df  Sum Sq  Mean Sq  F value    Pr(>F)
factor(treatment)  2   51.57   25.784    11.43  0.000963
Residuals         15   33.83   2.255
---                            ======
```

Summary of data:
$\bar{y}_{1\cdot} = 51.0, \bar{y}_{2\cdot} = 46.90, \bar{y}_{3\cdot} = 48.42, s_1 = 1.81, s_2 = 1.72, s_3 = 0.71,$
and $MSE = 2.26$.

# Pairwise comparisons

CI for $\mu_1 - \mu_2$: $\bar{y}_{1\cdot} - \bar{y}_{2\cdot} \pm t_{0.025,15}\sqrt{MSE}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} =$
$51 - 46.90 \pm 2.131\sqrt{2.26} * \sqrt{\frac{1}{6} + \frac{1}{6}} = 4.10 \pm 1.85 = (2.25, 5.95).$
Similarly, we can get
CI for $\mu_1 - \mu_3 : (0.73, 4.43)$
CI for $\mu_2 - \mu_3 : (-3.37, 0.33)$.

```
> qt(0.025,15)
[1] -2.13145

> qt(0.975,15)
[1] 2.13145
```

# Contrasts

A contrast is a linear combination of population means. It is a more general comparison of means.
contrast: $C = c_1\mu_1 + c_2\mu_2 + \cdot + c_t\mu_t$ where
$c_i$'s are constants such that $\sum c_i = 0$.

e.g., $C_1 = \mu_3 - \mu_4$
$C_2 = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \frac{1}{3}\mu_3 - \frac{1}{3}\mu_4 - \frac{1}{3}\mu_5$ are contrasts.
$C_2$ compares the average of $\mu_1$ and $\mu_2$ to the average of $\mu_3, \mu_4$ and $\mu_5$.

# Estimate $C$

$\hat{C} = c_1\bar{y}_{1\cdot} + \cdots + c_t\bar{y}_{t\cdot}$
If the t samples are random and independent samples from normal distributions with mean $\mu_i$ and common variance $\sigma^2$, then each $\bar{y}_{i\cdot} \sim N(\mu_i, \frac{\sigma^2}{n_i})$. As $\hat{C}$ is a linear combination of independent normal random variables, hence $\hat{C} \sim N(C, \sigma^2_{\hat{C}})$ with
$V(\hat{C}) = \sigma^2_{\hat{C}} = \sigma^2(\frac{c_1^2}{n_1} + \cdots + \frac{c_t^2}{n_t})$.
Fact: Suppose $X_1, X_2, \cdots, X_k$ are independent random variables, for constants $c_1, c_2, \cdots c_k$,
$\text{E}(c_1X_1 + c_2X_2 + \cdots + c_kX_k) = c_1\text{E}(X_1) + c_2\text{E}(X_2) + \cdots c_k\text{E}(X_k)$
and
$V(c_1X_1 + c_2X_2 + \cdots + c_kX_k) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots c_k^2V(X_k)$
An estimate of $\sigma^2_{\hat{C}}$ is $s^2_{\hat{C}} = \text{MSE}(\frac{c_1^2}{n_1} + \cdots + \frac{c_t^2}{n_t})$.

the standard error of $\hat{C}$ is $s_{\hat{C}} = \sqrt{s^2_{\hat{C}}}$.

# t test and CI

Fact: $t = \frac{\hat{C} - C}{s_{\hat{C}}} \sim t_\nu$ with $\nu = N - t$.
CI for C: $\hat{C} \pm t_{\alpha/2;N-t} s_{\hat{C}}$.
point estimate $\pm$ multiplier$\times$ standard error of point estimate

test $H_0 : C = 0, H_a : C \neq 0$.
test statistic $t = \frac{\hat{C}}{s_{\hat{C}}} \sim t_{N-t}$ under $H_0$.

# Estimate and test a contrast

Exercise: Use the data of example 5.1. Define contrast $C = \mu_1 - \frac{1}{2}\mu_2 - \frac{1}{2}\mu_3$ (compares mean 1 to the average of mean 2 and mean 3)
Test $H_0 : C = 0$ vs $H_a : C \neq 0$ at levele of significance $\alpha = 0.05$.
Also find a 95% CI for $C$.

# Solutions

Here $c_1 = 1, c_2 = -0.5, c_3 = -0.5$

$\hat{C} = \bar{y}_{1\cdot} - \frac{1}{2}\bar{y}_{2\cdot} - \frac{1}{2}\bar{y}_{3\cdot} = 51 - \frac{1}{2} * 46.90 - \frac{1}{2} * 48.42 = 3.34.$

$s^2_{\hat{C}} = 2.26 * (\frac{1^2}{6} + \frac{(-0.5)^2}{6} + \frac{(-0.5)^2}{6}) = 0.565, s_{\hat{C}} = 0.752.$

$t = \frac{3.34}{0.752} = 4.44.$
p-value $=2 * P(t > 4.44) = 0.0005$. Reject $H_0$.

CI for C is $3.34 \pm 2.131 * 0.752 = (1.74, 4.94)$.

```
> 2*pt(-4.44,15)
[1] 0.0004771517
```

# Exercise

Exercise 5.7.

```
Delay(min)  Angle(dgree)
 30        140,138,140,138,142
 45        140,150,120,128,130
 60        118,130,128,118,118
```

Perform an F test to examine if there is a difference in the mean angle among the three delay times.
If the test is significant (at $\alpha = 0.05$), get three CIs for $\mu_1 - \mu_2, \mu_1 - \mu_3$ and $\mu_2 - \mu_3$.
Also define a contrast $C = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3$ (this contrast compares the mean angle of short and medium delay to mean angle of long delay). Test $H_0 : C = 0$ vs $H_a : C \neq 0$ and obtain a 95% CI for $C$.

# Effect of multiple comparisons

- Overall or **experimentwise significance level** $\alpha_e$ : probability of making at least 1 type I error among $m$ tests. $\alpha_e \leq m\alpha$, where $\alpha$ is the significance level of each individual test.
- Overall **experimentwise confidence level** $CL_e$: probability that all confidence intervals are correct. $CL_e \geq 1 - m\alpha$, where $1 - \alpha$ is the confidence level of each individual CI.

  A CI is correct means it contains the true parameter it tries to estimate.

# Bonferroni method

Carry out each test at significance level $\frac{\alpha}{m}$ rather than $\alpha$ or multiply the p-value of each test by $m$ then compare to $\alpha$.
e.g., $m = 5$, if need $\alpha_e = 0.05$, reject each test if $p$-value$< 0.05/5 = 0.01$. or multiply each p-value by 5 then compare to 0.05.

In confidence interval, use critical value $t_{\alpha_e/(2m),\nu}$ rather than $t_{\alpha_e/2,\nu}$.
e.g., $m = 5$, if need $CL_e = 0.95$ (implying $\alpha_e = 0.05$), then use critical value in each CI $t_{0.025/5}$ instead of $t_{0.025}$.
In example 5.3. use critical value $t_{0.025/3,15} = 2.69$ for the three CIs as $m = 3$.
in R, it is easier to use left tail probability in qt function. $qt(\alpha_e/(2 * m), \nu)$ gives a negative critical value. Just drop the negative sign.

# R code for pairwise tests with Bonferroni adjustment

```
 g1 = c(9,12,10,8,15)
 g2 = c(20,21,23,17,30)
 g3 = c(6,5,8,16,7)
 y = c(g1,g2,g3)

type = c(rep(1,5),rep(2,5),rep(3,5))
type = factor (type)

pairwise.t.test(y,type,p.adj="none")
pairwise.t.test(y,type,p.adj="bonf")
```

# output

```
> pairwise.t.test(y,type,p.adj="none")
Pairwise comparisons using t tests with pooled SD
data:  y and type
  1       2
2 0.00089 -
3 0.37416 0.00019
P value adjustment method: none

>pairwise.t.test(y,type,p.adj="bonf")
Pairwise comparisons using t tests with pooled SD
data:  y and type
  1       2
2 0.00267 -
3 1.00000 0.00056
P value adjustment method: bonferroni
```

Note with Bonferroni adjustment, each p-value is multiplied by 3.

# Tukey-Cramer method

With Bonferroni adjustment, we use a larger t critical value in each CI and multiply the p-value of each test by $m$.
We can also use the quantiles and tail probabilities of the q distribution (studentized range distribution) to achieve similar results.
**Assume t independent random samples:**
$y_{11}, \cdots, y_{1n} \sim N(\mu_1, \sigma^2)$**,**
**....**
$y_{t1}, \cdots, y_{tn} \sim N(\mu_t, \sigma^2)$**.**
**Then under** $H_0 : \mu_1 = \mu_2 = \cdots = \mu_t$**,**
$q = \frac{\bar{y}_{max} - \bar{y}_{min}}{S_p \sqrt{\frac{1}{n}}}$ **(where** $S_p$ **is the pooled sample standard deviation) has a Studentized Range distribution with** $t$ **and** $\nu (\nu = nt - t$**,associated with** $S_P$**) degrees of freedom.**

# Tukey Test and CI

$H_0 : \mu_i = \mu_j; H_a : \mu_i \neq \mu_j$
test statistic value $q^* = \frac{|\bar{y}_{i.} - \bar{y}_{j.}|}{\sqrt{MSE}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}\sqrt{2}$,
p-value $= P(q > q^*)$

R code to get p-value: 1- ptukey($q^*$,t, $\nu$).

CI: Let the experimentwise CL is $1 - \alpha_e$.
The critical value is $q_{\alpha_e,t,\nu}/\sqrt{2}$ where $q_{\alpha_e}$ is quantile of the q distribution with upper tail probability $\alpha_e$, $\nu$ is the d.f. associated with MSE, and t is the number of groups to be compared.

$\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \pm \frac{q_{\alpha_e,t,\nu}}{\sqrt{2}}\sqrt{MSE}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$.

R code to get quantile of the q distribution $q_{0.05;3,15}$ :

```
> qtukey(0.95, 3, 15)
```

More general, qtukey(CLe, t, N-t).

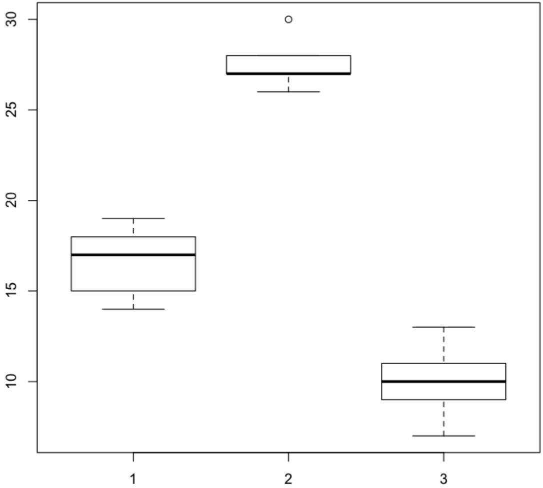
Problem 5.2:
time for ice cubes to melt in three beverages:

```
1. Coke    19,17, 15,14,18
2. Orange Juice 27,28, 30,26, 27
3. Water    10,11, 13,7,9
```

$\bar{y}_{1\cdot} = 16.6, \bar{y}_{2\cdot} = 27.6, \bar{y}_{3\cdot} = 10.0$
MSE=3.87 with 12 d.f. (F = 102.22, P-value $< 0.0001$).
Get Tukey CIs for $\mu_1 - \mu_2, \mu_1 - \mu_3, \mu_2 - \mu_3$ with experimentwise confidence level 99%.
What critical value would you use for Bonferroni CIs?

# Boxplot

```
   > boxplot(y~type)
or > boxplot(coke,juice,water)
```



# ANOVA table

```
> coke=c(19,17, 15,14,18)
> juice=c(27,28, 30,26, 27)
> water=c(10,11, 13,7,9)
> y=c(coke,juice,water)
> type=c(rep(1,5),rep(2,5),rep(3,5))
> type=factor(type)
> output=aov(y~type)
> anova(output)   #anova(output) can have higher precision
                  #than summary(output)

Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
type       2 790.53  395.27  102.22 2.904e-08 ***
Residuals 12  46.40    3.87
```

# Solutions

Note $\alpha = 0.01, t = 3, \nu = 12$.
so $q_{0.01,3,12}$=5.046, and the critical value is $5.046/\sqrt{2} = 3.568$.

```
> qtukey(0.99,3,12)
[1] 5.045934
```

The CIs are:
1 vs 2: $16.6 - 27.6 \pm 3.568 * \sqrt{3.87}\sqrt{\frac{1}{5} + \frac{1}{5}} = (-15.44, -6.56)$
1 vs 3: $16.6 - 10 \pm 3.568 * \sqrt{3.87}\sqrt{\frac{1}{5} + \frac{1}{5}} = (2.16, 11.04)$.
2 vs 3: $27.6 - 10 \pm 3.568 * \sqrt{3.87}\sqrt{\frac{1}{5} + \frac{1}{5}} = (13.16, 22.04)$.

The critical value for the Bonferroni CIs is 3.649.

```
> qt(0.01/6,12)
[1] -3.648889
```

Tukey adjustment is less conservative than Bonferroni adjustment when making all pairwise comparisons.

# R code for Tukey method

HSD: Honest Significant Difference

```
> output = aov(y~type)

> TukeyHSD(output,conf.level=.99)

  Tukey multiple comparisons of means
    99% family-wise confidence level

Fit: aov(formula = y ~ type)

$type
     diff        lwr        upr     p adj
2-1  11.0   6.562637  15.437363 0.0000037
3-1  -6.6 -11.037363  -2.162637 0.0005048
3-2 -17.6 -22.037363 -13.162637 0.0000000
```

# Tukey test

$H_0 : \mu_1 = \mu_3$
$H_a : \mu_1 \neq \mu_3$
$q^* = \frac{|10-16.6|}{\sqrt{3.87}\sqrt{1/5+1/5}} * \sqrt{2} = 7.502$
p-value= $P(q > 7.502) = 0.0005$ which matches the p-adj value in the R output.

```
  > 1-ptukey(7.502,3,12)
[1] 0.0005066512
```