

STAT 5309

LAB 3

****CONTENTS: Set up data frame- 1-Factor Design- Multiple comparisons-Contrasts- Power/sample size.**

***DUE: Sun, Feb 17**

A. PRACTICE

-----Balanced data; Linear model -----

Data: Bacteria under package methods. $\log(\text{count}/\text{cm}^2)$ on meat samples stored in 4 packaging conditions for 9 days.

Packaging Condition	$\log(\text{count}/\text{cm}^2)$
Commercial plastic wrap	7.66, 6.98, 7.80
Vacuum packaged	5.26, 5.44, 5.80,
1% CO ₂ , 40% O ₂ , 59% N	7.41, 7.33, 7.04
100% CO ₂	3.51, 2.91, 3.66

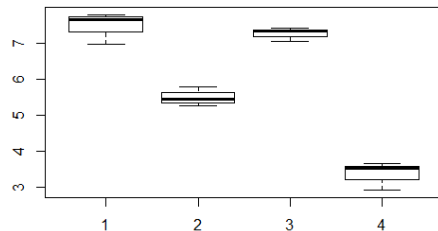
Note: N= 12 observations. Factor “package” has 4 treatment levels (a =4). Each treatment level has 3 replicates (n=3)

```
package <- rep( c(1,2,3,4) ,each=3))      # a=4 .
logcount <- c(7.66,6.98,7.80,5.26,5.44,5.80,7.41,7.33,7.04,3.51,2.91,3.66)
bacteria <- data.frame(package,logcount)
attach(bacteria)
package <- factor(package)
bacteria
```

	package	logcount
1	1	7.66
2	1	6.98
3	1	7.80
4	2	5.26
5	2	5.44
6	2	5.80
7	3	7.41
8	3	7.33
9	3	7.04
10	4	3.51
11	4	2.91
12	4	3.66

```
tapply(logcount,package,mean)    # treatment means, in a vector
tapply(logcount,package,sd)      #treatment standard deviations, in a vector

boxplot(logcount ~ package)      #Box Plot
```



```
>bact.mod <- aov(logcount ~ package)    #linear model
>summary.aov(bact.mod)                  #summary.aov(), anova(mod1) give same ANOVA
results
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
package	3	32.87	10.958	94.58	1.38e-06 ***
Residuals	8	0.93	0.116		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
>anova(bact.mod)                        #ANOVA analysis of model
>summary.lm(bact.mod)                   #summary with coefficients
```

```
Call:
aov(formula = logcount ~ package)

Residuals:
    Min       1Q   Median       3Q      Max
-0.500 -0.225  0.110  0.210  0.320

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.4800     0.1965  38.064 2.49e-10 ***
package2     -1.9800     0.2779  -7.125 9.95e-05 ***
package3     -0.2200     0.2779  -0.792  0.451
package4     -4.1200     0.2779 -14.825 4.22e-07 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3404 on 8 degrees of freedom
Multiple R-squared:  0.9726,    Adjusted R-squared:  0.9623
F-statistic: 94.58 on 3 and 8 DF,  p-value: 1.376e-06
```

Note: Control Level is Treatment #1. Treatment 1 mean is **7.4800**

```
>predict(m1)                                #prediction
```

1	2	3	4	5	6	7	8	9	10	11	12
7.48	7.48	7.48	5.50	5.50	5.50	7.26	7.26	7.26	3.36	3.36	3.36

```
##-----Model Without Intercept-----
```

```
>bact.mod1 <- aov(logcount ~ package -1)
```

```
>anova(bact.mod1)
```

Analysis of Variance Table

Response: logcount

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
package	4	450.59	112.648	972.36	8.861e-11 ***
Residuals	8	0.93	0.116		

```
>summary.lm(bact.mod1)
```

Call:

```
aov(formula = logcount ~ package - 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.500	-0.225	0.110	0.210	0.320

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
package1	7.4800	0.1965	38.06	2.49e-10 ***
package2	5.5000	0.1965	27.99	2.87e-09 ***
package3	7.2600	0.1965	36.94	3.16e-10 ***
package4	3.3600	0.1965	17.10	1.39e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3404 on 8 degrees of freedom

Multiple R-squared: 0.9979, Adjusted R-squared: 0.9969

F-statistic: 972.4 on 4 and 8 DF, p-value: 8.861e-11

#model matrices from 2 models

(Intercept) package1 package2 package3					package1 package2 package3 package4				
1	1	1	0	0	1	1	0	0	0
2	1	1	0	0	2	1	0	0	0
3	1	1	0	0	3	1	0	0	0

Notes:

- The coefficients are exactly the treatment means.
- Coefficients are **different** among 2 models
- Fitted and Prediction are the same.

```
> predict(bact.mod1)                                     #prediction
  1    2    3    4    5    6    7    8    9   10   11   12
7.48 7.48 7.48 5.50 5.50 5.50 7.26 7.26 7.26 3.36 3.36 3.36
```

##----- Multiple comparisons: TukeyHSD()-----

TukeyHSD(bact.mod)

```
Tukey multiple comparisons of means
 95% family-wise confidence level

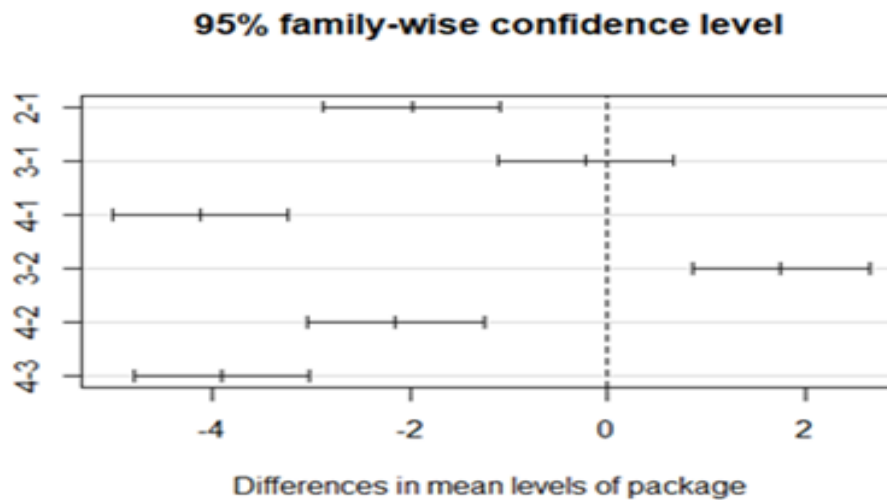
Fit: aov.default(formula = logcount ~ package)

$package
      diff      lwr      upr      p adj
2-1 -1.98 -2.869962 -1.090038 0.0004549
3-1 -0.22 -1.109962  0.669962 0.8563618
4-1 -4.12 -5.009962 -3.230038 0.0000020
3-2  1.76  0.870038  2.649962 0.0010160
4-2 -2.14 -3.029962 -1.250038 0.0002639
4-3 -3.90 -4.789962 -3.010038 0.0000031
```

Notes:

Interval (-1.109962 , 0.669962) **contain 0**. [Confirm with boxplot]. Also look at p-value.

```
## -----plot TukeyHSD results-----
Tukey <- TukeyHSD(mod1)
plot(Tukey)                                #Plot TukeyHSD
```



Note: 3-1 contains 0.

##-----**Contrasts: to compare one group of means to another group of means**-----

By default, R uses traditional dummy coding,

```
options("contrasts")    #to check what contrast R is using
```

\$contrasts	unordered	ordered
"contr.treatment"		"contr.poly"

```
options(contrasts= c("contr.treatment", "contr.poly")) #otherwise set back to default
```

R automatically uses the **first level** as the reference group. In some cases, there's another level you would prefer to use as the reference group, You can change the reference group with the `relevel()` command:

```
levels(package)
[1] "1" "2" "3" "4"

package <- relevel(package, ref="4")

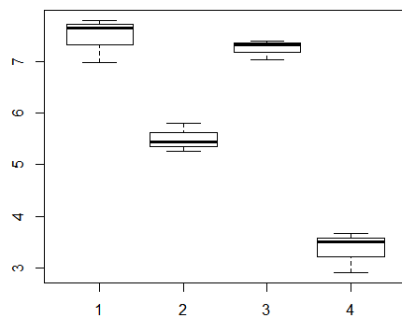
levels(package) # Note that the order of the levels has changed
[1] "4" "1" "2" "3"
class(package)
```

--	--

#-----**Build a contrast**-----

From the box plot, we notice there might be some relationships among means

Boxplot(logcount ~ package) # $\mu_1 = \mu_3$, $\mu_2 = \frac{1}{2}(\mu_3 + \mu_4)$



A Contrast L is a linear combination of the treatment means, to compare one group of means to another group of means. Testing a contrast is mainly testing against 0.

```
cont <- matrix(c(1,0,-1,0, 0,1, -1/2, -1/2), 4,2) # Form a matrix of contrasts , in columns
cont
```

```
> cont
      [,1] [,2]
[1,]     1  0.0
[2,]     0  1.0
[3,]    -1 -0.5
[4,]     0 -0.5
```

```
contrasts(package)<- cont
bact.mod2 <- aov(logcount ~ package, data=bacteria)
```

```
summary.aov(bact.mod2, split=list(package=list("Tr 1 is equal to Tr 3" = 1, "Tr 2 is average of Tr 3, 4"=2)))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
package	1	16.85	16.854	9.946	0.0103 *

```

package: Tr 1 is equal Tr 3      1 16.85 16.854 9.946 0.0103 *
package: Tr 2 is average of Tr 3, 4 1
Residuals                      10 16.95 1.695
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
#-----fit.contrast()-----
```

```
#Fit.contrast() [library "gmodels" ] test the contrast(s) .
```

```
library(gmodels)
```

```
cont1 <- t(cont)          #cont1 must have contrasts as ROW.
```

```
[ or  > cont1 <- matrix( c(1 ,0, -1, 0, 0, 1, -1/2, -1/2), 2,3, byrow=T]
```

```
rownames(cont1)<- c("Tr1 is equal to Tr 3", "Trt 2 is average of Tr3 and Tr4")
```

```
cont1
```

```
fit.contrast(bact.mod, "package", cont1, conf.int=.95)
```

	Estimate	Std. Error	t value	Pr(> t)
packageTr1 is equal to Tr 3	0.22	0.2779089	0.7916264	0.4514097
packageTrt 2 is average of Tr3 and Tr4	0.19	0.2406761	0.7894426	0.4526137
	lower CI	upper CI		
packageTr1 is equal to Tr 3	-0.4208590	0.8608590		
packageTrt 2 is average of Tr3 and Tr4	-0.3650002	0.7450002		

Note: p-values are large (compare to .05). Those relationships are true.

```
##-----Power of test-Sample size-----
```

```
# ----- power.anova.test() -----
```

```
# -----gives the power of the F-test in the model.
```

```
grp.means <- tapply(logcount, package, mean)
```

```
> power.anova.test(groups=4, n=3, between.var=var(grp.means), within.
var= 0.3^2, sig.level=.05, power=NULL)
```

Balanced one-way analysis of variance power calculation

```

groups = 4
n = 3
between.var = 3.652533
within.var = 0.09
sig.level = 0.05
power = 1

```

NOTE: n is number in each group

Note: Given n (as in the data set), power=NULL power.anova.test() gives the power.

Given groups=4, n=3, $\sigma = 0.3$ [from summary.aov()], power=1.[desirable power]
 Given power level, power.anova.test() gives the required number of replicates n.

-----Function Fpower1(), package “daewr”-----

```
library(daewr)
rmin <- 2
rmax <- 10
sigma <- 0.3 #given from summary.aov(mod1)
alpha <- .05 #given
Delta <- 1.7 # D= treatment means difference needs to be detected, from TukeyHSD
nlev <- 4 # 1 factor with 4 level
nreps <- c(rmin:rmax)
power <- Fpower1(alpha, nlev, nreps, Delta, sigma)
power
```

	alpha	nlev	nreps	Delta	sigma	power
[1,]	0.05	4	2	1.7	0.3	0.8218522
[2,]	0.05	4	3	1.7	0.3	0.9969953
[3,]	0.05	4	4	1.7	0.3	0.9999795
[4,]	0.05	4	5	1.7	0.3	0.9999999
[5,]	0.05	4	6	1.7	0.3	1.0000000
[6,]	0.05	4	7	1.7	0.3	1.0000000
[7,]	0.05	4	8	1.7	0.3	1.0000000
[8,]	0.05	4	9	1.7	0.3	1.0000000
[9,]	0.05	4	10	1.7	0.3	1.0000000

Note: n= 3 replicates is satisfactory number

B. EXERCISE

1. Data: Casting of High Temperature Alloys-

A metal alloy is produced by a high temperature casting process. The tensile strength of the alloy is critical for its future use. The casting process is designed produce bars of alloy with an average tensile strength above minimum requirement. An experiment was planned to isolate the variation in tensile strength due to the effects of different castings. 30 bars or alloy were produced using the 3 casting methods.

	Casting
1	88.0, 88.0, 94.8, 90.8, 93.0, 89.0, 86.0, 92.9, 89.0, 93.0
2	85.9, 88.6, 90.0, 87.1, 85.6, 86.0, 91.0, 89.6, 93.0, 87.5
3	94.2, 91.5, 92.0, 96.5, 95.6, 93.8, 92.5, 93.2, 96.2, 92.5

- (a) Set up the data frame. There are 30 observations.
Form a factor vector “casting”. Form a vector of response “strength”
Form a data named “alloy”.
- (b) Find the treatment means, treatment vars, treatment standard deviations
- (c) Do a boxplot, stripchart.
- (d) Build a linear model, using aov(). Do a summary.lm() and a summary.aov()
- (e) Perform a Pairwise.t.test(p.adj=”bonf”)
- (f) Perform TukeyHSD(). Plot the TukeyHSD result.
- (g) Calculate the power of the F-test in the aov model.

2. Data: Detection of Phlebitis on Rabbits.

Rabbits are used as the test animals, randomly assigned to 3 treatment groups. Intravenous needles are inserted in a vein of one ear. Three (3) intravenous treatments are administered to test animals. (a) Amiodarone with a vehicle solution (b) Vehicle solution alone and (c) Saline solution. .

An increase in temperature of the two ears (treated minus untreated) was used as the response

	Temp difference
Amiodarone	2.2, 1.6, 0.8, 1.8, 1.4, 0.4, 0.6, 1.5, 0.5
Vehicle	0.3, 0.0, 0.6, 0.0, -0.3, 0.2
Saline	0.1, 0.1, 0.2, -0.4, 0.3, 0.1, 0.1, -0.5

(a) Set up the data frame. There are 23 observations. Create a vector “trt” of 3 levels
 Hint: : `trt <- c(rep(1,time=9),rep(2, time=6),rep(3,time=8))`. A vector “tempdiff” of response.
 Create a data frame named :rabbits.

- (b) Perform treatment means, treatment variances, treatment standard deviations
- (c) Boxplot. Strip chart.
- (d) Build a linear model, using `aov()`. Summary
- (e) Perform pairwise t -test.
- (f) Perform TukeyHSD. Plot the TukeyHSD result.

3.

3-10. The response time in milliseconds was determined for three different types of circuits that could be used in an automatic valve shutoff mechanism. The results are shown in the following table:

Circuit Type		Response Time			
1	9	12	10	8	15
2	20	21	23	17	30
3	6	5	8	16	7

- (a) Test the hypothesis that the three circuit types have the same response time. Use $\alpha = 0.01$.
- (b) Use Tukey’s test to compare pairs of treatment means. Use $\alpha = 0.01$.
- (c) Use the graphical procedure in Section 3-5.3 to compare the treatment means. What conclusions can you draw? How do they compare with the conclusions from part (b)?
- (d) Construct a set of orthogonal contrasts, assuming that at the outset of the experiment you suspected the response time of circuit type 2 to be different from the other two.
- (e) If you were the design engineer and you wished to minimize the response time, which circuit type would you select?
- (f) Analyze the residuals from this experiment. Are the basic analysis of variance assumptions satisfied?

--