

HW #3

(7.10.1)

$$z_0 = z_{(1-0.05)} = 1.645; \quad z_1 = z_{(1-0.1)} = 1.282$$

$$n = \left(\frac{(1.645 + 1.282)(32)}{516 - 620} \right)^2 = 547.56 \simeq 548$$

$$C = 516 + 1.645 \frac{32}{\sqrt{548}} = 518.25; \quad C = 520 - 1.28 \frac{32}{\sqrt{548}} = 518.25$$

Select a sample of size 548 and compute \bar{x} . If $\bar{x} \geq 518.25$, reject H_0 . If $\bar{x} < 518.25$, do not reject H_0 .

(7.10.2)

$$z_0 = z_{(1-0.01)} = 2.326; \quad z_1 = z_{(1-0.05)} = 1.645$$

$$n = \left(\frac{(2.326 + 1.645)(0.02)}{4.500 - 4.520} \right)^2 = 15.8 \simeq 16$$

$$C = 4.500 + 2.326 \frac{0.02}{\sqrt{16}} = 4.512; \quad C = 4.520 - 1.645 \frac{0.02}{\sqrt{16}} = 4.512$$

Select a sample of size 16 and compute \bar{x} . If $\bar{x} \geq 4.512$, reject H_0 . If $\bar{x} < 4.512$, do not reject H_0 .

Ex 8.2.1

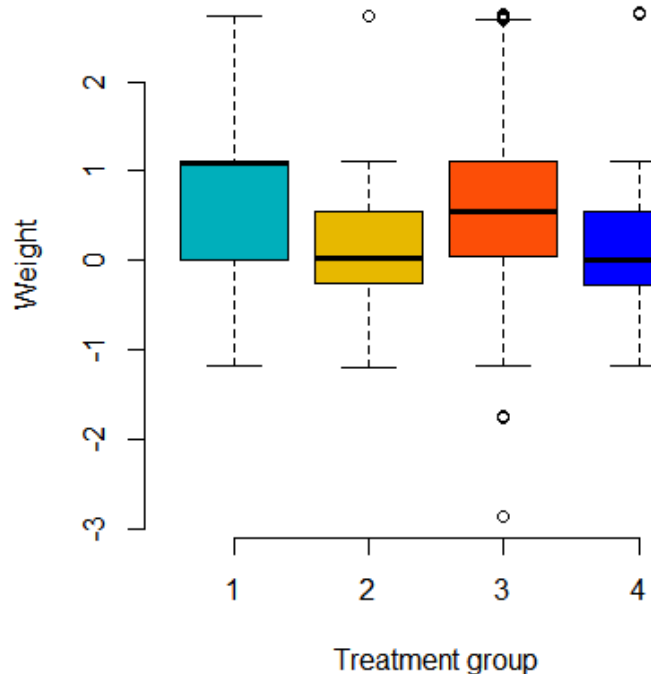
```
> hw821 <- read.csv("C:/Users/dlabate/Desktop/Teaching/ma4310/EXR_C08_S02_01.csv")
> str(hw821)
```

```
'data.frame':  329 obs. of  2 variables:
 $ weight: num  -0.003 0.05 0.272 0.552 1.116 ...
 $ group : int   1 1 1 1 1 1 1 1 1 1 ...
```

Note that the group variable is not recognized as a factor. Hence we need to reassign it.

```
> hw821$group <- factor(hw821$group, levels = c(1,2,3,4),labels = c("a30", "a60", "a90", "a120"))
> str(hw821)
'data.frame':  329 obs. of  2 variables:
 $ weight: num  -0.003 0.05 0.272 0.552 1.116 ...
 $ group : Factor w/ 4 levels "a30","a60","a90",...: 1 1 1 1 1 1 1 1 1 1 ...
```

```
> boxplot(weight ~ group, data = hw821, xlab = "Treatment group", ylab = "Weight",
+         frame = FALSE, col = c("#00AFBB", "#E7B800", "#FC4E07", "#0000FF"))
```



```
> res.aov <- aov(weight ~ group, data = hw821)
> summary(res.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	3	20.0	6.669	6.242	0.000393 ***
Residuals	325	347.2	1.068		

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

```
> qf(0.95,df1=3,df2=325)
```

```
[1] 2.632394
```

Conclusion: Since $F = 6.242 > 2.737492$ or since $p < 0.001$, we reject H_0 .

Weight differed significantly among the four training types.

Next we run the Tukey HSD test to determine which pairs of groups are statistically different.

```
> TukeyHSD(res.aov)
```

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

```
Fit: aov(formula = weight ~ group, data = hw821)
```

```
$group
```

	diff	lwr	upr	p adj
a60-a30	-0.447587615	-1.0125162	0.1173410	0.1734399
a90-a30	0.191059839	-0.1830744	0.5651941	0.5516339
a120-a30	-0.443557312	-1.0084859	0.1213713	0.1799776
a90-a60	0.638647454	0.1360387	1.1412562	0.0062693
a120-a60	0.004030303	-0.6530720	0.6611326	0.9999986
a120-a90	-0.634617151	-1.1372259	-0.1320084	0.0067138

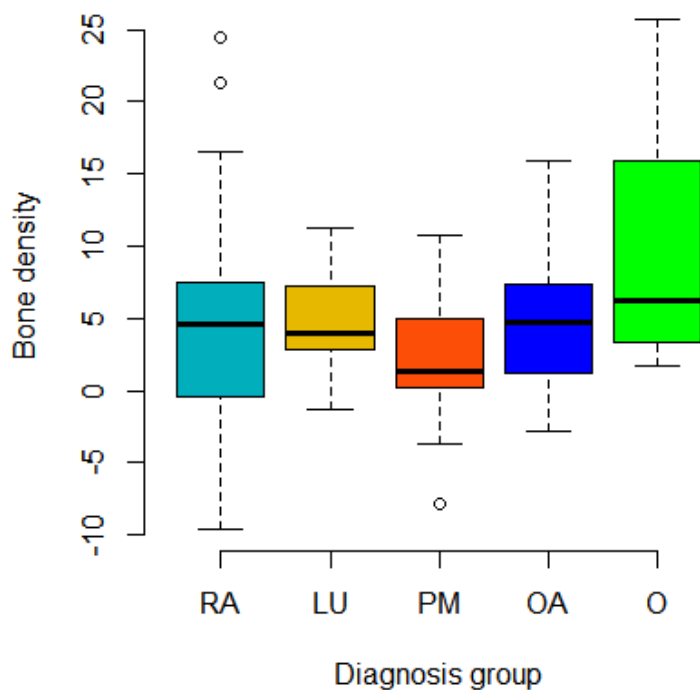
Conclusion: Only the differences a90-a60 and a120-a90 are significant

Ex 8.2.2

```
> hw822 <- read.csv("C:/Users/dlabate/Desktop/Teaching/ma4310/EXR_C08_S02_02.csv")
> str(hw822)
'data.frame': 96 obs. of 2 variables:
 $ weight: num 11.09 24.41 10.03 -3.16 6.83 ...
 $ group : int 1 1 1 1 1 1 1 1 1 1 ...
```

Note that the group variable is not recognized as a factor. Hence we need to reassign it.

```
> hw822$group <- factor(hw822$group, levels = c(1,2,3,4,5), labels = c("RA",
"LU", "PM", "OA", "O"))
> str(hw822)
'data.frame': 96 obs. of 2 variables:
 $ weight: num 11.09 24.41 10.03 -3.16 6.83 ...
 $ group : Factor w/ 5 levels "RA","LU","PM",...: 1 1 1 1 1 1 1 1 1 1 ...
> boxplot(weight ~ group, data = hw822, xlab = "Diagnosis group", ylab = "Bone density", frame = FALSE, col = c("#00AFBB", "#E7B800", "#FC4E07", "#0000FF", "#00FF00"))
```



```
> res.aov <- aov(weight ~ group, data = hw822)
> summary(res.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	4	355	88.86	2.277	0.067 .
Residuals	91	3551	39.02		

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

Conclusion: Since $p > 0.05$, we accept H_0 .

Bone density did not differ significantly among the five diagnosis groups. Clearly, no Tukey HSD test is needed in this case.

QUIZ #3 A study investigated the dietary intake of calcium among a cross section of 113 healthy women ages 20–88.

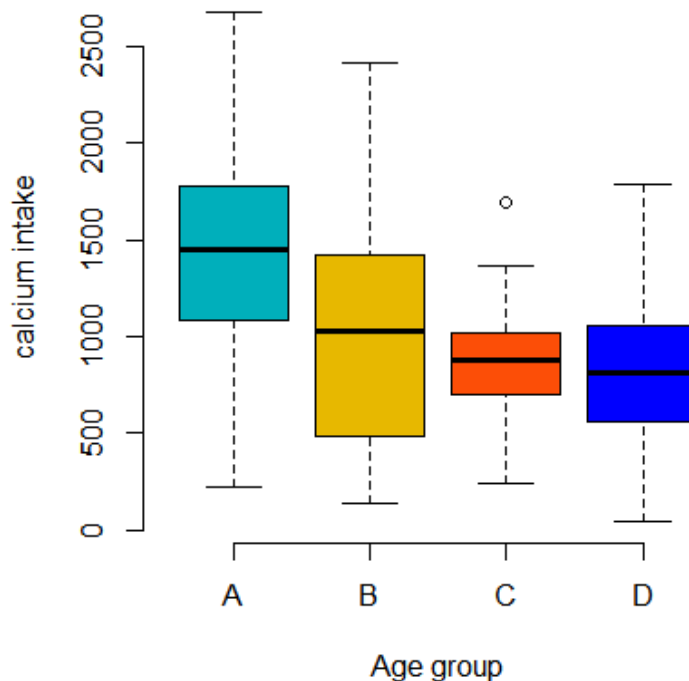
The researchers formed four age groupings as follows: Group A, 20.0–45.9 years; group B, 46.0–55.9 years; group C, 56.0–65.9 years; and group D, over 66 years.

Calcium from food intake was measured in mg/day.

- 1) Generate side-by-side boxplots of the data
- 2) Apply the analysis of variance hypothesis testing to see if you can conclude that there is a difference among population means. Justify your conclusion.
- 3) Use the Tukey's HSD procedure to test for significant differences among individual pairs of means, if appropriate. Justify your conclusion.
- 4) Apply the Shapiro-Wilk test to check that the normality assumption of the data is satisfied.

```
> dataq3 <- read.csv("C:/Users/dlabate/Desktop/Teaching/ma4310/dataq3.csv")
> str(dataq3)
'data.frame': 113 obs. of 2 variables:
 $ calcium: int 1820 2588 2670 1022 1555 222 1197 1249 1520 489 ...
 $ Group : chr "A" "A" "A" "A" ...
> dataq3$Group <- factor(dataq3$Group, levels = c("A","B","C","D"),labels = c
("A", "B", "C", "D"))
> str(dataq3)
'data.frame': 113 obs. of 2 variables:
 $ calcium: int 1820 2588 2670 1022 1555 222 1197 1249 1520 489 ...
 $ Group : Factor w/ 4 levels "A","B","C","D": 1 1 1 1 1 1 1 1 1 1 ...
 1. Generate side-by-side boxplots of the data
```

```
> boxplot(calcium ~ Group, data = dataq3, xlab = "Age group", ylab = "calcium
intake",frame = FALSE, col = c("#00AFBB", "#E7B800", "#FC4E07", "#0000FF"))
```



2. Apply the analysis of variance hypothesis testing

```
> res.aov <- aov(calcium ~ Group, data = dataq3)
> summary(res.aov)
          Df    Sum Sq Mean Sq F value    Pr(>F)
Group      3  5931208 1977069    9.359 1.48e-05 ***
Residuals 109 23026500  211252
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: since $p\text{-value} < 0.05$, we reject H_0 and conclude that there is a statistically significant difference among the means.

3. Use the Tukey's HSD procedure

```
> TukeyHSD(res.aov)
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = calcium ~ Group, data = dataq3)

$Group
      diff      lwr      upr      p adj
B-A -455.72078 -865.7066  -45.73492 0.0230454
C-A -574.53605 -913.5892 -235.48294 0.0001356
D-A -596.63447 -905.3867 -287.88228 0.0000109
C-B -118.81527 -509.0844  271.45386 0.8568796
D-B -140.91369 -505.1676  223.34020 0.7443051
D-C  -22.09842 -304.1436  259.94679 0.9969598
```

Conclusion: based on the values of p adj, only the differences between the groups A-B, A-C and A-D are statistically significant.

4. Apply the Shapiro-Wilk test

```
> aov_residuals <- residuals(object = res.aov)
> shapiro.test(x = aov_residuals)

      Shapiro-Wilk normality test

data:  aov_residuals
W = 0.97799, p-value = 0.05908
```

Conclusion: since the $p\text{-value} > 0.05$, we can accept H_0 that data are approximately normal. In this case, the normality assumption is barely met.