

Test #1

Please, write clearly and justify your work to receive credit. Whenever you use R, you must report the R command you entered with the complete list of parameters. You also need to report the R output (numerical p-value, for instance) that you used to draw your conclusion.

(1) [5Pts] *The following data are the oxygen uptakes (milliliters) during incubation of a random sample of 13 cell suspensions:*

14.1, 14.5, 13.2, 11.2, 14.0, 12.2, 11.1, 13.7, 13.2, 16.0, 12.8, 14.4, 12.9

(a) *Use an appropriate statistical test to check whether data can be assumed to satisfy the normality assumption.*

(b) *Do these data provide sufficient evidence at the 0.01 level of significance that the population mean is not 12 ml? You must state the hypothesis testing problem you are solving and justify your conclusion.*

(a)

```
> x=c(14.1, 14.5, 13.2, 11.2, 14.0, 12.2, 11.1, 13.7, 13.2, 16.0, 12.8, 14.4, 12.9)
> shapiro.test(x)
```

Shapiro-Wilk normality test

data: x

W = 0.96688, p-value = 0.8544

Conclusion: Since the p-value > 0.05, we can assume that data are normally distributed.

(b) We test $H_0 : \bar{x} = 12$ against $H_1 : \bar{x} \neq 12$ with $\alpha = 0.01$.

From the data: $\bar{x} = 13.331$, $s^2 = 1.842$, $s = 1.357$, $n = 13$.

Test statistic (Student t pdf):

$$t = \frac{\bar{x}_r - x_0}{\sqrt{\frac{s^2}{n}}} = \frac{13.331 - 12.000}{\sqrt{\frac{1.842}{13}}} = 3.535$$

Rejection region: $t > t_{0.01/2;12} \Rightarrow qt(1 - 0.01/2, 12) = 3.055$

Conclusion: since $t > t_{0.01/2;12}$, then H_0 is REJECTED.

R solution

```
> t.test(x,mu=12,alternative="two.sided" )
```

One Sample t-test

data: x

t = 3.535, df = 12, p-value = 0.004108

alternative hypothesis: true mean is not equal to 12

Conclusion: Since p-value is less than 0.01, then H_0 is REJECTED.

(2) [4Pts] *A study about the effects of reminiscence therapy for older women with depression considers a sample of 10 women residing in an assisted living long-term care facility. For this study, depression was measured by the Geriatric Depression Scale (GDS). Higher scores indicate more severe depression symptoms. The participants received reminiscence therapy. Pre-treatment and posttreatment depression scores are given in the following table.*

Pre-GDS: 12, 10, 16, 2, 12, 18, 11, 16, 16, 10

Post-GDS: 11, 10, 11, 3, 9, 13, 8, 14, 14, 10

Can we conclude that subjects who participate in reminiscence therapy experience, on average, a decline in GDS depression scores? Let $\alpha = 0.01$. You must state the hypothesis testing problem you are solving and justify your conclusion.

We apply a Paired t-test. We set $d = x_{pre} - x_{post}$.
 We test $H_0 : \mu_d \leq 0$ against $H_1 : \mu_d > 0$ with $\alpha = 0.01$.
 Data: $n = 10$, $\bar{d} = \frac{1}{10} \sum_{i=1}^{15} d_i = 2$, $s_d = 2.055$.
 Test statistic (Student t pdf):

$$t = \frac{\bar{d} - \mu_d}{s_{\bar{d}}} = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \frac{2}{\frac{2.055}{\sqrt{10}}} = 3.078$$

Rejection region: $t > t_{0.005;9} = 2.821$
 Since $t > t_{0.005;9}$, then H_0 is REJECTED.

R solution

```
> x=c(12, 10, 16, 2, 12, 18, 11, 16, 16, 10)
> y=c(11, 10, 11, 3,9, 13, 8, 14, 14, 10)
> t.test(x,y,alternative = "greater", paired = TRUE,var.equal = TRUE)
```

Paired t-test

data: x and y

t = 3.0779, df = 9, p-value = 0.006592

alternative hypothesis: true mean difference is greater than 0

Conclusion: Since p-value is less than 0.01, then H_0 is REJECTED.

(3) [7Pts] A study of Marfan syndrome (a genetic disorder that affects connective tissue) reported the following severity scores of patients with no dural ectasia (NO), mild dural ectasia (MI) and marked dural ectasia (MA):

NO : 18, 18, 20, 21, 23, 23, 24, 26, 26, 27, 28, 20, 29, 21, 20, 21, 20, 18

MI : 10, 16, 22, 22, 23, 26, 28, 28, 28, 29, 29, 30, 31, 32, 32, 33, 33, 38, 39, 40, 47

MA : 17, 24, 26, 27, 29, 30, 30, 33, 34, 35, 35, 36, 39

Note that data is available on file: test1.csv

(a) May we conclude, on the basis of these data, that mean severity scores differ among the three populations represented in the study? Let $\alpha = 0.05$.

(b) Use Tukey's procedure to test for significant differences among individual pairs of sample means.

[Extra credit +2] (c) Verify that normality of the data and homogeneity of the variance are satisfied using appropriate statistical tests.

```
> severity <-c(18, 18, 20, 21, 23, 23, 24, 26, 26, 27, 28, 20, 29, 21, 20, 21, 20,
18, 10, 16, 22, 22, 23, 26, 28, 28, 28, 29, 29, 30, 31, 32, 32, 33, 33,38, 39, 40,
47,17, 24, 26, 27, 29, 30, 30, 33, 34, 35, 35, 36, 39)
> group <-factor(c(rep("NO",len=18),rep("MI",len=21),rep("MA",len=13)))
> Data <- data.frame(severity, group)
> str(Data)
'data.frame': 52 obs. of 2 variables:
$ severity: num 18 18 20 21 23 23 24 26 26 27 ...
```

```
$ group : Factor w/ 3 levels "MA","MI","NO": 3 3 3 3 3 3 3 3 3 3 ...
```

Alternatively, if data are loaded from the file test1.csv

```
> Data <- read.csv("C:/Users/dlabate/Desktop/Teaching/ma4310/test1.csv")
> Data$group <- factor(Data$group, levels = c(1,2,3),labels = c("NO", "MI","MA"))
> res.aov <- aov(severity ~ group, data = Data)
> str(test1)
'data.frame': 52 obs. of 2 variables:
 $ severity: int 18 18 20 21 23 23 24 26 26 27 ...
 $ group : Factor w/ 3 levels "NO","MI","MA": 1 1 1 1 1 1 1 1 1 1 ...
```

(a)

```
> res.aov <- aov(severity ~ group, data = Data)
> summary(res.aov)
```

	<i>Df</i>	<i>SumSq</i>	<i>MeanSq</i>	<i>Fvalue</i>	<i>Pr(> F)</i>	
<i>group</i>	2	644.1	322.0	7.937	0.00103	**
<i>Residuals</i>	49	1988.0	40.6			

Conclusion: Since p-value is less than 0.05, then H_0 is REJECTED, that is, mean severity scores differ among the three populations.

(b)

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

Fit: aov(formula = severity ~ group, data = Data)

<i>\$group</i>		<i>diff</i>	<i>lwr</i>	<i>upr</i>	<i>padj</i>
<i>MI - MA</i>	-1.051282	-6.484201	4.381637	0.8867111	
<i>NO - MA</i>	-7.995726	-13.599080	-2.392373	0.0032883	
<i>NO - MI</i>	-6.944444	-11.889390	-1.999499	0.0038543	

Conclusion: only the differences NO-MA and NO-MI are significant but the difference MI-MA is not.

(c)

We test normality first:

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

Shapiro-Wilk normality test

data: aov_residuals

W = 0.96338, p-value = 0.1097

Since the p-value is above 0.05, we can assume normality of the data.

Finally we test homogeneity.

```
> library(car)
> leveneTest(severity ~ group, data = Data)
```

Levene's Test for Homogeneity of Variance (center = median)

	<i>Df</i>	<i>Fvalue</i>	<i>Pr(> F)</i>
<i>group</i>	2	2.4294	0.09863
	49		

Since the p-value is above 0.05, can assume the homogeneity of variances.