

HW5

Jinhong Du, 15338039

Content

Pulmonary Disease	2
12.6	2
12.7	4
12.8	5
Gastroenterology	5
12.42	6
12.43	6

Pulmonary Disease

Twenty-two young asthmatic volunteers were studied to assess the short-term effects of sulfur dioxide (SO_2) exposure under various conditions [12]. The baseline data in Table 12.30 were presented regarding bronchial reactivity to SO_2 stratified by lung function (as defined by forced expiratory volume / forced vital capacity [FEV_1/FVC]) at screening.

Lung-function group		
Group B $FEV_1 / FVC \leq 74\%$	Group A $FEV_1 / FVC 75 - 84\%$	Group C $FEV_1 / FVC \geq 85\%$
20.8	7.5	9.2
4.1	7.5	2.0
30.0	11.9	2.5
24.7	4.5	6.1
13.8	3.1	7.5
	8.0	
	4.7	
	28.1	
	10.3	
	10.0	
	5.1	
	2.2	

Table 12.30 Relationship of bronchial reactivity to SO_2 (cm H_2O/s) grouped by lung function at screening among 22 asthmatic volunteers

12.6 Test the hypothesis that there is an overall mean difference in bronchial reactivity among the three lung-function groups.

Suppose that

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2)$$

To test

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0 \quad H_1 : \exists i \in \{1, 2, 3\} \text{ s.t. } \alpha_i \neq 0$$

```
Y1 <- c(20.8,4.1,30.0,24.7,13.8)
Y2 <- c(7.5,7.5,11.9,4.5,3.1,8.0,4.7,28.1,10.3,10.0,5.1,2.2)
Y3 <- c(9.2,2.0,2.5,6.1,7.5)

n1 <- length(Y1)
```

```

n2 <- length(Y2)
n3 <- length(Y3)

df1 <- 3 - 1
df2 <- n1 + n2 + n3 - 3

Y_mean <- (sum(Y1) + sum(Y2) + sum(Y3))/(n1+n2+n3)
Y1_mean <- mean(Y1)
Y2_mean <- mean(Y2)
Y3_mean <- mean(Y3)

# F test
MSB <- (n1*(Y1_mean - Y_mean)^2 + n2*(Y2_mean - Y_mean)^2 + n3*(Y3_mean - Y_mean)^2) / df1
MSW <- (sum((Y1 - Y1_mean)^2) + sum((Y2 - Y2_mean)^2) + sum((Y3 - Y3_mean)^2)) / df2
F_ <- MSB / MSW
cat('The value of the F statistics is ', F_, '\n',
    'F(0.95,', 3-1, ', ', n1 + n2 + n3 - 3, ') is ', qf(0.95, df1, df2), '\n',
    'p-value is ', 1 - pf(F_, df1, df2), '\n')

## The value of the F statistics is 4.989255
## F(0.95, 2 , 19 ) is 3.521893
## p-value is 0.01813163

# With build-in function
data=rbind(
data.frame(name="Y1" ,value=Y1),
data.frame(name="Y2" ,value=Y2),
data.frame(name="Y3" ,value=Y3)
);
fit <- aov(value~name,data=data);
summary(fit)

##              Df Sum Sq Mean Sq F value Pr(>F)
## name          2   503.5   251.77    4.989 0.0181 *
## Residuals    19   958.8    50.46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Since p -value is less than 0.05, we reject H_0 , i.e., there is an overall mean difference in bronchial reactivity among the three lung-function groups.

12.7 Compare the means of each pair of groups using the LSD method.

```
data <- data.frame(
  value= c(Y1,Y2,Y3),
  name=factor(c(rep(1,n1),rep(2,n2),rep(3,n3)))
)

t_value <- c()
p_value <- c()
for (i in 1:2){
  for (j in (i+1):3){
    y1 <- data[data$name==i,1]
    y2 <- data[data$name==j,1]
    t_ <- (mean(y1) - mean(y2))/sqrt(MSW*(1/length(y1)+1/length(y2)))
    t_value <- c(t_value, t_)
    p_value <- c(p_value, 2 * (1 - pt(abs(t_), df2)))
  }
}

# With build-in function for computing p-value
pairwise.t.test(data$value, data$name, p.adjust.method = "none")

##
## Pairwise comparisons using t tests with pooled SD
##
## data:  data$value and data$name
##
##      1      2
## 2 0.0151 -
## 3 0.0084 0.4203
##
## P value adjustment method: none
```

The result is shown in [Table 2 Result](#). Only the p -value for the t test between group 2 and 3 is not significant.

12.8 Compare the means of each pair of groups using the Bonferroni method.

```
alpha_star <- 0.05 / choose(3,2)
cat('t(',df2,',',1 - alpha_star/2,') = ', qt(1-alpha_star/2,df2))

## t( 19 , 0.9916667 ) = 2.625106
```

Compare the absolute values of t -values with $t(19, 1 - \frac{1}{2}\alpha^*)$, we have

Groups	t -value	p -value	LSD	Bonferroni
1,2	2.6723892	0.0150588	H_1	H_1
1,3	2.9424811	0.0083594	H_1	H_1
2,3	0.8237993	0.4202747	H_0	H_0

Table 2: Result

Gastroenterology

In Table 12.39, we present data relating protein concentration to pancreatic function as measured by trypsin secretion among patients with cystic fibrosis [16].

Trypsin secretion [U/(kg/hr)]					
≤ 50		51 – 1000		> 10000	
Subject number	Protein concentration	Subject number	Protein concentration	Subject number	Protein concentration
1	17	1	1.4	1	1.7
2	2.0	2	2.4	2	2.0
3	2.0	3	2.4	3	2.0
4	2.2	4	3.3	4	2.2
5	4.0	5	4.4	5	4.0
6	4.0	6	4.7	6	4.0
7	5.0	7	6.7	7	5.0
8	6.7	8	7.6	8	6.7
9	7.8	9	9.5	9	7.8
		10	11.7		

Trypsin secretion [U/(kg/hr)]					
≤ 50		51 – 1000		> 10000	
Subject	Protein	Subject	Protein	Subject	Protein
number	concentration	number	concentration	number	concentration

Table 12.39 Relationship between protein concentration (mg/mL) of duodenal secretions to pancreatic function as measured by trypsin secretion [U/(kg/hr)]

12.42 If we do not want to assume normality for these distributions, then what statistical procedure can be used to compare the three groups?

We can use non-parametric test - the Kruskal-Wallis test.

12.43 Perform the test mentioned in Problem 12.42, and report a p-value. How do your results compare with a parametric analysis of the data?

```
Y1 <- c(1.7,2.0,2.0,2.2,4.0,4.0,5.0,6.7,7.8)
Y2 <- c(1.4,2.4,2.4,3.3,4.4,4.7,6.7,7.6,9.5,11.7)
Y3 <- c(2.9,3.8,4.4,4.7,5.0,5.6,7.4,9.4,10.3)
n1 <- length(Y1)
n2 <- length(Y2)
n3 <- length(Y3)
data <- data.frame(
  value= c(Y1,Y2,Y3),
  name=factor(c(rep(1,n1),rep(2,n2),rep(3,n3)))
)

kruskal.test(data$value, data$name)

##
## Kruskal-Wallis rank sum test
##
## data:  data$value and data$name
```

```
## Kruskal-Wallis chi-squared = 2.8618, df = 2, p-value = 0.2391
```

```
fit <- aov(value~name,data=data)
summary(fit)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## name       2  19.63   9.814   1.265   0.3
## Residuals 25 193.99   7.760
```

The p -value of the Kruskal-Wallis test is less significant than the one of one-way ANOVA. Both tests accept H_0 .