

# **BMEG 802 – Advanced Biomedical Experimental Design and Analysis**

## Assignment 4

---

Joshua G. A. Cashaback, PhD

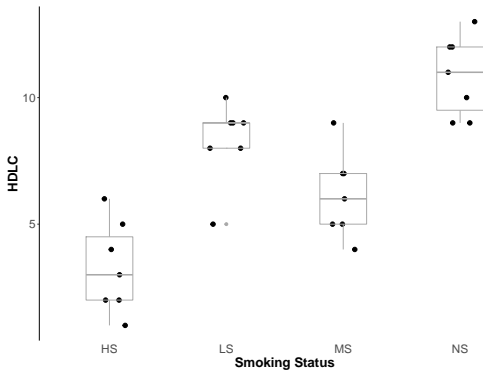
# Question 1

```
install.packages("ggplot2")

hdlc <- c(12, 10, 11, 13, 9, 9, 12, 9, 8, 5, 9, 9, 10, 8, 5, 4, 7, 9, 5, 7, 6, 3,
smoker <- c("NS","NS","NS","NS","NS","NS","NS","NS","LS","LS","LS","LS","LS","LS","LS","LS"
library(ggplot2)
# create data frame
data1 <- data.frame(
  SMOKER=smoker,
  HDLC=hdlc,
  sd=c(1) # dummy input
)
```

# Graphing with GG Plot - Box and Whisker

```
ggplot(data1, mapping=aes(x=SMOKER, y=HDL)) +  
  geom_point(position=position_jitter(width=0.15, height=0), color="black", size=2.5) +  
  geom_boxplot(fill=NA, width=0.4, size=0.5, color="DarkGray") +  
  theme_classic() +  
  theme(axis.title=element_text(size=18,face="bold"), axis.text=element_text(size=16),  
        axis.line.x=element_blank(), axis.ticks.x=element_blank()) +  
  labs(x="Smoking Status", y="HDL")
```



## Question 1 Cont'd

```
HDLC <- c(12, 10, 11, 13, 9, 9, 12, 9, 8, 5, 9, 9, 10, 8, 5, 4, 7, 9, 5, 7)
Smoker <- c(1,1,1,1,1,1,1,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,4,4,4,4,4,4,4)
m1 <- lm(HDLC ~ factor(Smoker))
anova(m1)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: HDLC
```

```
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## factor(Smoker)  3 216.857   72.286    26.06 1.005e-07 ***
## Residuals      24  66.571    2.774
```

```
## ---
```

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

# Question 1 Cont'd

```
Group1 = hdlc[1:7]  
Group2 = hdlc[8:14]  
Group3 = hdlc[15:21]  
Group4 = hdlc[22:28]  
shapiro.test(Group1)$p.value
```

```
## [1] 0.4196491
```

```
shapiro.test(Group2)$p.value
```

```
## [1] 0.06607612
```

```
shapiro.test(Group3)$p.value
```

```
## [1] 0.7385166
```

```
shapiro.test(Group4)$p.value
```

```
## [1] 0.7724094
```

Since  $p > 0.05$  for all, no violations of normality

## Question 1 Cont'd

```
bartlett.test(HDLC ~ factor(Smoker))
```

```
##  
## Bartlett test of homogeneity of variances  
##  
## data: HDLC by factor(Smoker)  
## Bartlett's K-squared = 0.12176, df = 3, p-value = 0.9891
```

Since  $p > 0.05$ , no violation of homogeneity of variances

## Question 1 Cont'd

```
install.packages("effectsize")
```

```
library(effectsize)
```

```
omega_squared(m1)
```

```
## For one-way between subjects designs, partial omega squared is equivalent to
```

```
## Returning omega squared.
```

```
## Parameter      | Omega2 |      90% CI
```

```
## -----
```

```
## factor(Smoker) |    0.73 | [0.54, 0.82]
```

# Question 1 Cont'd

```
options(scipen = 999) # shows all values in decimal form
pval_1v2 = t.test(Group1, Group2, alternative = "two.sided")$p.value # G1 vs G2
pval_1v3 = t.test(Group1, Group3, alternative = "two.sided")$p.value # G1 vs G3
pval_1v4 = t.test(Group1, Group4, alternative = "two.sided")$p.value # G1 vs G4
pval_2v3 = t.test(Group2, Group3, alternative = "two.sided")$p.value # G2 vs G3
pval_2v4 = t.test(Group2, Group4, alternative = "two.sided")$p.value # G2 vs G4
pval_3v4 = t.test(Group3, Group4, alternative = "two.sided")$p.value # G3 vs G4
pvals = c(pval_1v2, pval_1v3, pval_1v4, pval_2v3, pval_2v4, pval_3v4)
pvals

## [1] 0.010506033434 0.000155856817 0.000002627869 0.030962641135 0.000145470449
## [6] 0.009697653922

p.adjust(pvals, method = "holm", n = length(pvals))

## [1] 0.02909296177 0.00072735225 0.00001576721 0.03096264113 0.00072735225
## [6] 0.02909296177
```



## Question 1 Cont'd

```
install.packages("effsize")
```

```
library(effsize)
```

```
options(scipen = 999) # shows all values in decimal form
```

```
d_1v2 = cohen.d(Group1, Group2, var.equal = False)$estimate # G1 vs G2
```

```
d_1v3 = cohen.d(Group1, Group3, var.equal = False)$estimate # G1 vs G3
```

```
d_1v4 = cohen.d(Group1, Group4, var.equal = False)$estimate # G1 vs G4
```

```
d_2v3 = cohen.d(Group2, Group3, var.equal = False)$estimate # G2 vs G3
```

```
d_2v4 = cohen.d(Group2, Group4, var.equal = False)$estimate # G2 vs G4
```

```
d_3v4 = cohen.d(Group3, Group4, var.equal = False)$estimate # G3 vs G4
```

```
d_effsize = c(d_1v2, d_1v3, d_1v4, d_2v3, d_2v4, d_3v4)
```

```
d_effsize
```

```
## [1] 1.618625 2.899874 4.479318 1.306407 2.933693 1.643065
```

## Question 1 Cont'd

There was a significant main effect of smoking status on HDLC [ $F(3,24) = 26.06$ ,  $p < 0.001$ ,  $\omega^2 = 0.73$ ]. All groups were significantly different from one another ( $p < 0.05$  and  $d > 1.31$  for all mean comparisons). It is worth noting that the light smokers had significantly higher HDLC compared to moderate smokers.

## Question 1 Cont'd

```
kruskal.test(HDLC ~ factor(Smoker))
```

```
##  
##  Kruskal-Wallis rank sum test  
##  
## data:  HDLC by factor(Smoker)  
## Kruskal-Wallis chi-squared = 20.744, df = 3, p-value = 0.000119
```

## Question 2

```
power.anova.test(groups=4, n=NULL, between.var=var(c(79,82,81,88)),  
                  within.var=10**2, sig.level=0.05, power=0.80)
```

```
##  
##      Balanced one-way analysis of variance power calculation  
##  
##      groups = 4  
##      n = 25.22645  
##      between.var = 15  
##      within.var = 100  
##      sig.level = 0.05  
##      power = 0.8  
##  
## NOTE: n is number in each group
```

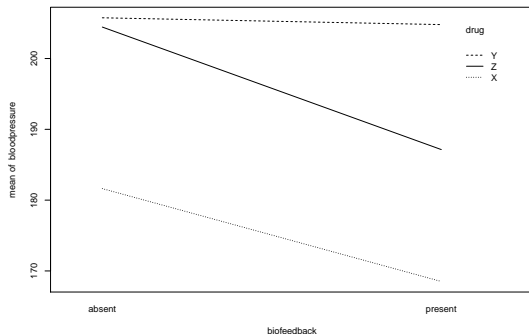
We should collect 26 participants per group to be sufficiently powered.

## Question 3

```
dx_bf <- rnorm(20, mean = 170, sd=12.91)
dy_bf <- rnorm(20, mean = 203, sd=13.91)
dz_bf <- rnorm(20, mean = 188, sd=13.84)
dx_nbf <- rnorm(20, mean = 186, sd=11.84)
dy_nbf <- rnorm(20, mean = 201, sd=11.93)
dz_nbf <- rnorm(20, mean = 210, sd=12.81)
```

## Question 3 Cont'd

```
bloodpressure <- c(dx_bf, dy_bf, dz_bf, dx_nbf, dy_nbf, dz_nbf)
biofeedback <- factor(c(rep("present",60),rep("absent",60)))
drug <- factor(rep(c(rep("X",20),rep("Y",20),rep("Z",20)),2))
bpdata <- data.frame(bloodpressure, biofeedback, drug)
interaction.plot(biofeedback, drug, bloodpressure)
```



## Question 3 Cont'd

```
myanova <- aov(bloodpressure ~ biofeedback*drug)
summary(myanova)
```

##	Df	Sum Sq	Mean Sq	F value	Pr(>F)						
## biofeedback	1	3282	3282	21.830	0.00000821 ***						
## drug	2	19087	9544	63.476	< 0.000000000000000002 ***						
## biofeedback:drug	2	1441	720	4.792	0.01 *						
## Residuals	114	17140	150								
## ---											
## Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'.'	0.1	' '	1

## Question 3 Cont'd

```
shapiro.test(bloodpressure[1:20])$p.value # group 1
```

```
## [1] 0.2561596
```

```
shapiro.test(bloodpressure[21:40])$p.value # group 2
```

```
## [1] 0.0679432
```

```
shapiro.test(bloodpressure[41:60])$p.value # group 3
```

```
## [1] 0.2260702
```

```
shapiro.test(bloodpressure[61:80])$p.value # group 4
```

```
## [1] 0.9518
```

```
shapiro.test(bloodpressure[81:100])$p.value # group 5
```

```
## [1] 0.5419854
```

```
shapiro.test(bloodpressure[101:120])$p.value # group 6
```

```
## [1] 0.8039848
```

No violations of normality



## Question 3 Cont'd

Test for sphericity (homogeneous of variance) in each main effect and interaction

```
bartlett.test(bloodpressure ~ interaction(drug, biofeedback))$p.value
```

```
## [1] 0.2097777
```

```
bartlett.test(bloodpressure ~ drug)$p.value
```

```
## [1] 0.8007082
```

```
bartlett.test(bloodpressure ~ biofeedback)$p.value
```

```
## [1] 0.05050119
```

## Question 3 Cont'd

```
install.packages("effectsize")
```

```
library(effectsize)
```

```
omega_squared(myanova)
```

## Parameter	Omega2 (partial)	90% CI
## -----		
## biofeedback	0.15	[0.06, 0.25]
## drug	0.51	[0.40, 0.59]
## biofeedback:drug	0.06	[0.00, 0.13]

## Question 3 Cont'd

```
# biofeedback
pval_dxbf_v_dybf = t.test(dx_bf, dy_bf, alternative = "two.sided")$p.value
pval_dxbf_v_dzbf = t.test(dx_bf, dz_bf, alternative = "two.sided")$p.value
pval_dybf_v_dzbf = t.test(dy_bf, dz_bf, alternative = "two.sided")$p.value

# no biofeedback
pval_dxnbf_v_dynbf = t.test(dx_nbf, dy_nbf, alternative = "two.sided")$p.value
pval_dxnbf_v_dznbf = t.test(dx_nbf, dz_nbf, alternative = "two.sided")$p.value
pval_dynbf_v_dznbf = t.test(dy_nbf, dz_nbf, alternative = "two.sided")$p.value
pvals = c(pval_dxbf_v_dybf, pval_dxbf_v_dzbf, pval_dybf_v_dzbf, pval_dxnbf_v_dynbf, pval_dxnbf_v_dznbf, pval_dynbf_v_dznbf)
pvals

## [1] 0.0000000005284255 0.0000309517829200 0.0005426306515590 0.0000001150953406
## [5] 0.0000000782763442 0.6864762759436223

p.adjust(pvals, method = "holm", n = length(pvals))

## [1] 0.000000003170553 0.000092855348760 0.001085261303118 0.000000460381362
## [5] 0.000000391381721 0.686476275943622
```

## Question 3 Cont'd

```
install.packages("effsize")
```

```
library(effsize)
options(scipen = 999) # shows all values in decimal form
d1v2 = cohen.d(dx_bf, dy_bf, var.equal = False)$estimate #
d1v3 = cohen.d(dx_bf, dz_bf, var.equal = False)$estimate #
d2v3 = cohen.d(dy_bf, dz_bf, var.equal = False)$estimate #
d4v5 = cohen.d(dx_nbf, dy_nbf, var.equal = False)$estimate #
d4v6 = cohen.d(dx_nbf, dz_nbf, var.equal = False)$estimate #
d5v6 = cohen.d(dy_nbf, dz_nbf, var.equal = False)$estimate #
d_effsize1 = c(d1v2, d1v3, d2v3, d4v5, d4v6, d5v6)
abs(d_effsize1)
```

```
## [1] 2.7230155 1.5089173 1.1961034 2.0602965 2.1383831 0.1286744
```

## Question 3 Cont'd

There was a significant interaction between drug and biofeedback on blood pressure ( $p < 0.001, \omega_p^2 = 0.11$ ). All mean comparisons showed that Drugs X, Y, and Z were significantly different from each other both in the presence of biofeedback ( $p < 0.05, d > 0.9$  for all comparisons) and the absence of biofeedback ( $p < 0.05, d > 0.72$  for all comparisons)

## Question 3 Cont'd

```
install.packages("pwr")
```

```
library(pwr)  
# u = levels, v = dof residuals (df_residuals = a * b * (n-1))  
pwr.f2.test(u=2, v=NULL, f2=0.1^2, sig.level=0.05, power=0.90)
```

```
##  
##      Multiple regression power calculation  
##  
##           u = 2  
##           v = 1265.395  
##           f2 = 0.01  
##      sig.level = 0.05  
##           power = 0.9
```

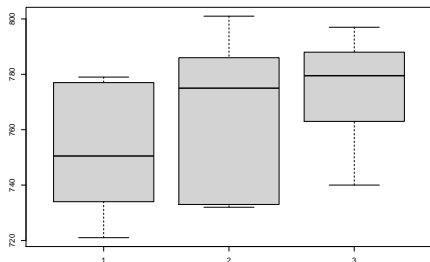
```
1265.395 / (3 * 2) + 1
```

```
## [1] 211.8992
```

We would need to collect 212 participants.

## Question 4.

```
dv <- c(745,764,774,777,786,788,734,733,763,779,801,797,756,786,  
       785,721,732,740)  
treatment <- factor(rep(c(1,2,3),6))  
subject <- factor(c(rep(1,3),rep(2,3),rep(3,3),  
                   rep(4,3),rep(5,3),rep(6,3)))  
mydata <- data.frame(dv, treatment, subject)  
boxplot(split(mydata$dv,mydata$treatment))
```



## Question 4 Cont'd

```
m1 <- aov(dv ~ treatment + Error(subject/treatment), data=mydata)
summary(m1)
```

```
##
## Error: subject
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  5   8548    1710
##
## Error: subject:treatment
##           Df Sum Sq Mean Sq F value  Pr(>F)
## treatment  2 1575.0   787.5   14.43 0.00113 **
## Residuals 10   545.7    54.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



## Question 4 Cont'd

```
install.packages("car")
```

```
response <- with(mydata, cbind(dv[treatment==1], dv[treatment==2], dv[treatment==3]))
```

```
mlm1 <- lm(response ~ 1)
```

```
rfactor <- factor(c("r1", "r2", "r3"))
```

```
design = data.frame(rfactor)
```

```
library(car)
```

```
## Loading required package: carData
```

```
mlm1.aov <- Anova(mlm1, idata=design, idesign = ~rfactor, type="III")
```

## Question 4 Cont'd

```
summary(mlm1.aov, multivariate=FALSE)
```

```
## Warning in summary.Anova.mlm(mlm1.aov, multivariate = FALSE): HF eps > 1 treated
## as 1
```

```
##
## Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
##
##           Sum Sq num Df Error SS den Df  F value    Pr(>F)
## (Intercept) 10520285      1   8547.8      5 6153.773 0.000000006378 ***
## rfactor      1575      2    545.7     10  14.432    0.001128 **
## ---
```

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

```
##
```

```
##
```

```
## Mauchly Tests for Sphericity
```

```
##
```

```
##           Test statistic p-value
```

```
## rfactor      0.75802 0.57459
```

```
##
```

```
##
```

```
## Greenhouse-Geisser and Huynh-Feldt Corrections
```

```
## for Departure from Sphericity
```

```
##
```

```
##           GG eps Pr(>F[GG])
```

```
## rfactor 0.80516    0.00286 **
```

```
## ---
```

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

## Question 4 Cont'd

```
pval = 0.00286
GGeps = 0.80516
df1_adj = 2 * GGeps # adjusted df
df2_adj = 10 * GGeps
Fscore = qf(1 - pval, df1=df1_adj, df2=df2_adj) # chi-square function
df1_adj
```

```
## [1] 1.61032
```

```
df2_adj
```

```
## [1] 8.0516
```

```
Fscore
```

```
## [1] 14.43117
```

# Question 4 Cont'd

```
summary(mlm1.aov, multivariate=TRUE)
```

```
## Warning in summary.Anova.mlm(mlm1.aov, multivariate = TRUE): HF eps > 1 treated
## as 1

##
## Type III Repeated Measures MANOVA Tests:
##
## -----
##
## Term: (Intercept)
##
## Response transformation matrix:
##      (Intercept)
## [1,]           1
## [2,]           1
## [3,]           1
##
## Sum of squares and products for the hypothesis:
##      (Intercept)
## (Intercept)    31560854
##
## Multivariate Tests: (Intercept)
##      Df test stat approx F num Df den Df      Pr(>F)
## Pillai      1    0.9992 6153.773      1      5 0.0000000063781 ***
## Wilks       1    0.0008 6153.773      1      5 0.0000000063781 ***
## Hotelling-Lawley 1 1230.7545 6153.773      1      5 0.0000000063781 ***
## Roy         1 1230.7545 6153.773      1      5 0.0000000063781 ***
```

## Question 4 Cont'd

```
install.packages("rstatix")
```

```
library(rstatix)
```

```
##
```

```
## Attaching package: 'rstatix'
```

```
## The following objects are masked from 'package:effectsize':
```

```
##
```

```
##      cohens_d, eta_squared
```

```
## The following object is masked from 'package:stats':
```

```
##
```

```
##      filter
```

```
friedman_test(dv ~ treatment | subject, data=mydata)
```

```
## # A tibble: 1 x 6
```

```
##   .y.      n statistic    df      p method
```

```
## * <chr> <int>      <dbl> <dbl> <dbl> <chr>
```

```
## 1 dv          6         7     2 0.0302 Friedman test
```

## Question 4 Cont'd

Sphericity not violated (Mauchly Test:  $p = 0.575$ ). Main Effect of Time is significant without ( $p = 0.001$ ) and with the Greenhouse-Geisser Corrections ( $p = 0.003$ ), with the Multivariate Approach ( $p = 0.007$ ), and Friedman Test ( $p = 0.030$ ).

## Question 4 Cont'd

Let's quickly split our data into each treatment

```
treat1 = response[1:6]  
treat2 = response[7:12]  
treat3 = response[13:18]
```

## Question 4 Cont'd

Performed on the PAIRED differences

```
shapiro.test(treat2-treat1)$p.value #
```

```
## [1] 0.9786552
```

```
shapiro.test(treat3-treat1)$p.value #
```

```
## [1] 0.1051113
```

```
shapiro.test(treat3-treat2)$p.value #
```

```
## [1] 0.2199981
```

Normality is not Violated, since all  $p > 0.05$



## Question 4 Cont'd

```
library(effectsize)
omega_squared(mlm1.aov)
```

```
## Parameter | Omega2 (partial) |      90% CI
## -----
## rfactor   |                0.86 | [0.29, 0.94]
```

## Question 4 Cont'd

```
p1v2 = t.test(treat1, treat2, paired = TRUE, alternative = "two.sided")$p.value
p1v3 = t.test(treat1, treat3, paired = TRUE, alternative = "two.sided")$p.value
p2v3 = t.test(treat2, treat3, paired = TRUE, alternative = "two.sided")$p.value
pvals = c(p1v2, p1v3, p2v3)
pvals_holm = p.adjust(pvals, method = "holm", n = length(pvals))
sprintf("%.5f", pvals_holm) #outputs in decimal (not scientific notation)
```

```
## [1] "0.04023" "0.00239" "0.19328"
```

## Question 4 Cont'd

```
library(effsize)  
cohen.d(treat1, treat2, paired = TRUE)$estimate #
```

```
## [1] -0.4824522
```

```
cohen.d(treat1, treat3, paired = TRUE)$estimate #
```

```
## [1] -0.9637995
```

```
cohen.d(treat2, treat3, paired = TRUE)$estimate #
```

```
## [1] -0.2168753
```

## Question 4 Cont'd

There was a significant main of effect of Time ( $F(1.6,8.1) = 14.4$ ,  $p = 0.003$ ,  $\omega^2 = 0.86$ ). Time 2 ( $p = 0.040$ ,  $d = 0.48$ ) and Time 3 ( $p = 0.002$ ,  $d = 0.96$ ) were significantly greater than Time 1. Thus, this medication actually led to more coughs and can be considered ineffective.

Note: interpretation from any omnibus test is acceptable.

## Question 4 Cont'd

```
library(WebPower)
```

```
## Loading required package: MASS
```

```
##
```

```
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:rstatix':
```

```
##
```

```
##      select
```

```
## Loading required package: lme4
```

```
## Loading required package: Matrix
```

```
## Registered S3 methods overwritten by 'lme4':
```

```
##   method                                from
```

```
##   cooks.distance.influence.merMod car
```

```
##   influence.merMod                    car
```

```
##   dfbeta.influence.merMod             car
```

```
##   dfbetas.influence.merMod            car
```

```
## Loading required package: lavaan
```

```
## This is lavaan 0.6-7
```

```
## lavaan is BETA software! Please report any bugs.
```

```
## Loading required package: parallel
```

```
## Loading required package: PearsonDS
```

## Question 4 Cont'd

```
# n=sub, ng=#ofgroups, nm=#ofmeasurements, nscor=sphericity(1=perfect)  
#type "0" between-effect; "1" within-effect; and "2" interaction effect  
wp.rmanova(n = NULL, ng = 1, nm = 3, f = .4, nscor = 0.8,  
           alpha = 0.05, power = 0.8, type = 1)
```

```
## Repeated-measures ANOVA analysis
```

```
##
```

```
##           n    f ng nm nscor alpha power
```

```
##       72.05928 0.4  1  3   0.8  0.05   0.8
```

```
##
```

```
## NOTE: Power analysis for within-effect test
```

```
## URL: http://psychstat.org/rmanova
```

73 Participants would be needed.