**Regression Lab 1**

---

The data set cholesterol.txt available on your thumb drive contains the following variables:

**Field Descriptions**

`ID`: Subject ID

`sex`: Sex: 0 = male, 1 = female

`age`: Age in years

`chol`: Serum total cholesterol, mg/dl

`BMI`: Body-mass index, $kg/m^2$

`TG`: Serum triglycerides, mg/dl

`APOE`: Apolipoprotein E genotype, with six genotypes coded 1-6: 1 = e2/e2, 2 = e2/e3, 3 = e2/e4, 4 = e3/e3, 5 = e3/e4, 6 = e4/e4

`rs174548`: Candidate SNP 1 genotype, chromosome 11, physical position 61,327,924. Coded as the number of minor alleles: 0 = C/C, 1 = C/G, 2 = G/G.

`rs4775401`: Candidate SNP 2 genotype, chromosome 15, physical position 59,476,915. Coded as the number of minor alleles: 0 = C/C, 1 = C/T, 2 = T/T.

The goal of the regression labs will be to use the data set to explore the relationship between triglycerides and several predictor variables. The objective of this first lab will be

- Become familiar with R and RStudio
- Begin to explore the cholesterol dataset.
- Use graphical methods to investigate associations between triglycerides and BMI

1. Open RStudio.

2. Create a new script file to record your R code. Open a script file by clicking on File -> New File -> R Script (for Mac).

3. Load the cholesterol data set.

4. Compute the sample mean, median and standard deviation of triglycerides.

5. View the boxplot, stem-and-leaf displays and histograms for triglycerides.

6. Create a variable called IBMI that takes the value 1 if BMI > 25 and 0 if BMI<=25 .

7. Compute summary measures of triglycerides for the two groups of subjects defined by IBMI.

8. Plot boxplots for triglycerides separately for the two groups of subjects defined by IBMI. Does there appear to be an association between BMI and triglycerides? Conduct a test of the null hypothesis that mean triglycerides do not differ between those with BMI > 25 and BMI <=25.

9. Plot a scatterplot of triglycerides vs BMI. Based on this plot does there appear to be an association between BMI and triglycerides? What can you additionally say about the relationship between these variables that was not possible using the boxplot?

10. Use regression to investigate the association between triglycerides and BMI. What do the linear regression model results tell us about the association?

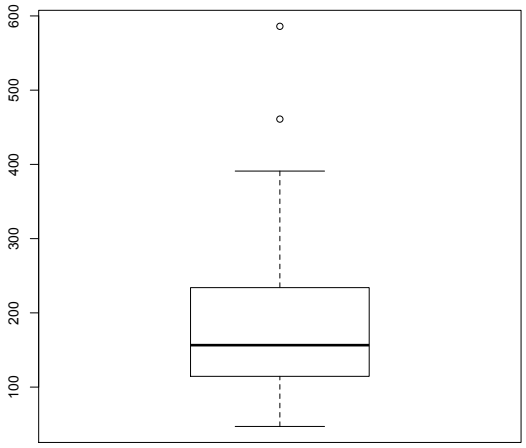
11. Compute the predicted value and its 95% confidence interval for the mean value of triglycerides at BMI = 23 as well as for a new individual with BMI = 23. How do these two intervals differ and why?

12. Check your script file. Make sure that all important commands that you have used and any output you want to save are included in here.

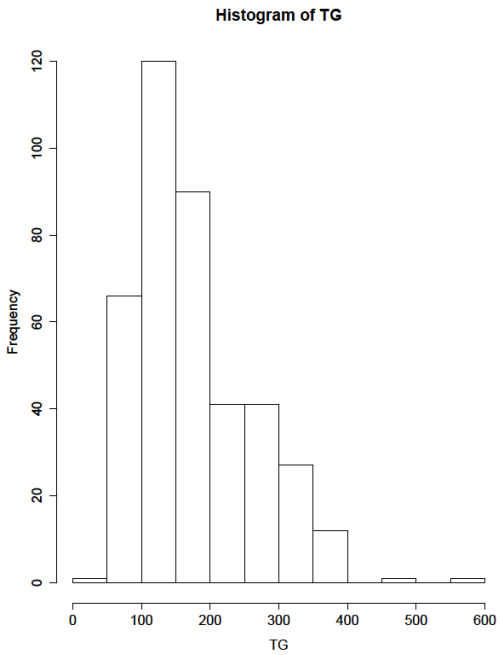
---

**R Commands & Output:**

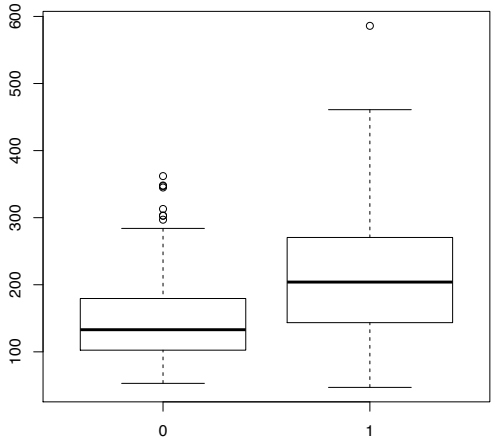
```
> cholesterol = read.table("http://faculty.washington.edu/rhubb/sisg/SISG-Data-
cholesterol.txt", header=T)
> attach(cholesterol)
>
> # compute univariate summary statistics for triglycerides
> mean(TG)
[1] 177.44
> median(TG)
[1] 156.5
> sd(TG)
[1] 82.98323
> summary(TG)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   47.0   114.8   156.5   177.4   234.0   586.0
>
> # graphical displays for triglycerides
> boxplot(TG)
```



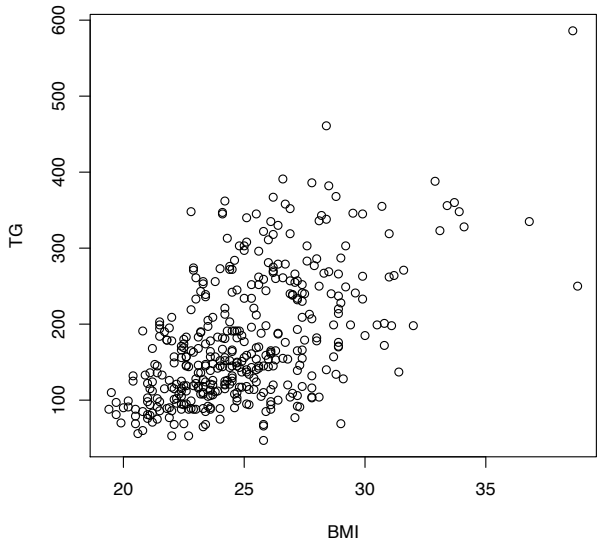

```
> stem(TG)

  The decimal point is 2 digit(s) to the right of the |

  0 |
  0 | 55566777777777888888888899999999999999999999999999999
  1 | 000000000000000000000011111111111111111111111222222222222222222222222+34
  1 | 555555555555555555555555556666666666666666666677777777777777777778888+18
  2 | 000000000000011111111122223333333344444444444444
  2 | 55555555666666666666777777777778888888899
  3 | 0000001112222233444444
  3 | 5555555566666778999
  4 |
  4 | 6
  5 |
  5 | 9

> hist(TG)
>
```




```
> # create a binary indicator for BMI > 25
> ibmi = ifelse(BMI > 25, 1, 0)
>
> # compute univariate summary statistics for triglycerides for BMI > 25 and BMI <= 25
> tapply(TG,ibmi,mean)
       0        1
147.3839 215.6932
> tapply(TG,ibmi,median)
  0   1
133 204
> tapply(TG,ibmi,sd)
       0        1
61.70787 90.66584
>
> # plot boxplots for triglycerides separately by BMI > 25 and BMI <= 25
> boxplot(TG ~ ibmi)
```



```
> t.test(TG ~ ibmi)

                              Welch Two Sample t-test

data:  TG by ibmi
t = -8.5584, df = 294.91, p-value = 6.391e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -84.01732 -52.60118
sample estimates:
mean in group 0 mean in group 1
       147.3839        215.6932

> # scatterplot of triglycerides vs BMI
> plot(BMI, TG)
```




```
> # fit linear regression models for the association between triglycerides and BMI
> fit1 = lm(TG ~ BMI)
> summary(fit1)

Call:
lm(formula = TG ~ BMI)

Residuals:
    Min      1Q  Median      3Q     Max
-170.19  -45.10  -12.89   39.60  231.08

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -208.50      28.95  -7.203 2.97e-12 ***
BMI            15.44       1.15  13.429  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 68.93 on 398 degrees of freedom
Multiple R-Squared: 0.3118, Adjusted R-squared: 0.3101
F-statistic: 180.3 on 1 and 398 DF,  p-value: < 2.2e-16

> predict(fit1, newdata = data.frame(BMI = 23), interval = "confidence")
       fit      lwr      upr
1 146.5612 138.4161 154.7062
> predict(fit1, newdata = data.frame(BMI = 23), interval = "prediction")
       fit      lwr      upr
1 146.5612 10.80972 282.3126
```

**Regression Lab 2**

---

The goal of this lab is to answer the following scientific questions using the cholesterol dataset.

- Are triglyceride levels associated with BMI?
- Are linear regression model assumptions satisfied for this relationship?
- Is the association between triglyceride and BMI modified by the APOE4 allele?

1) Load the `gee` package.

2) Construct a scatterplot of triglycerides versus BMI. Are there any points that you suspect might have a large influence on the regression estimates?

3) Use regression to investigate the association between triglycerides and BMI.after removing the observations with BMI > 37. Do the points with BMI>37 appear to affect your results? How?

4) Use residuals analysis to check the linear regression model assumptions. Create a scatterplot of residuals vs fitted values and a quantile-quantile plot of residuals. Do any modeling assumptions appear to be violated? How do model results change if you use robust standard errors?

5) Investigate the association between triglycerides and BMI after log transforming triglycerides. Does this appear to correct violations of modeling assumptions?

6) Create a new binary variable indicating presence of the APOE4 allele (APOE = 3, 5, or 6).

7) Plot separate scatterplots for triglycerides vs BMI for subjects in the two groups defined by presence of the APOE4 allele. Do these plots suggest effect modification?

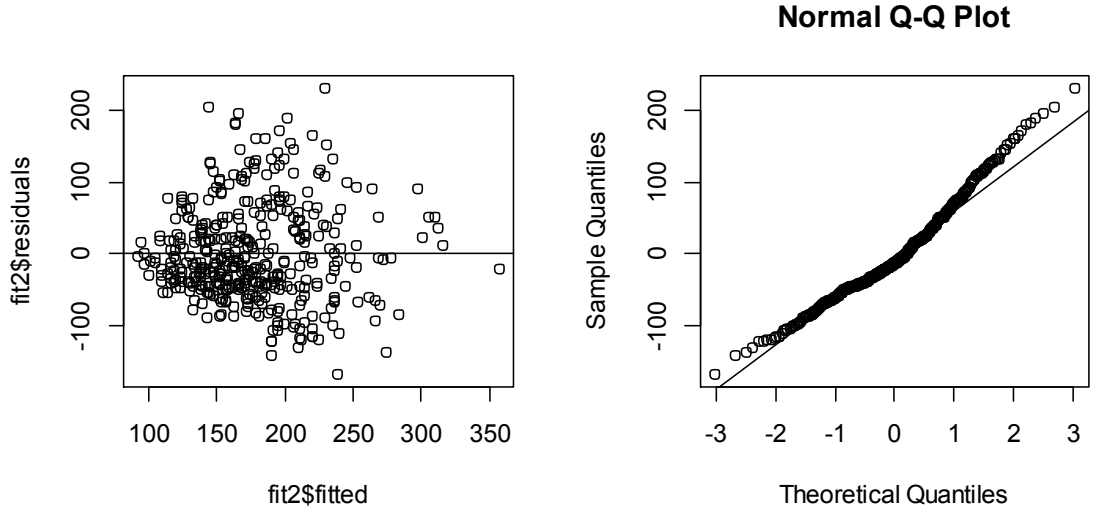
8) Fit a linear regression model that investigates whether the association between triglycerides and BMI is modified by the APOE4 allele. Is there an association between APOE4 and triglycerides? Is there evidence of effects modification?

---

**R Commands & Output:**

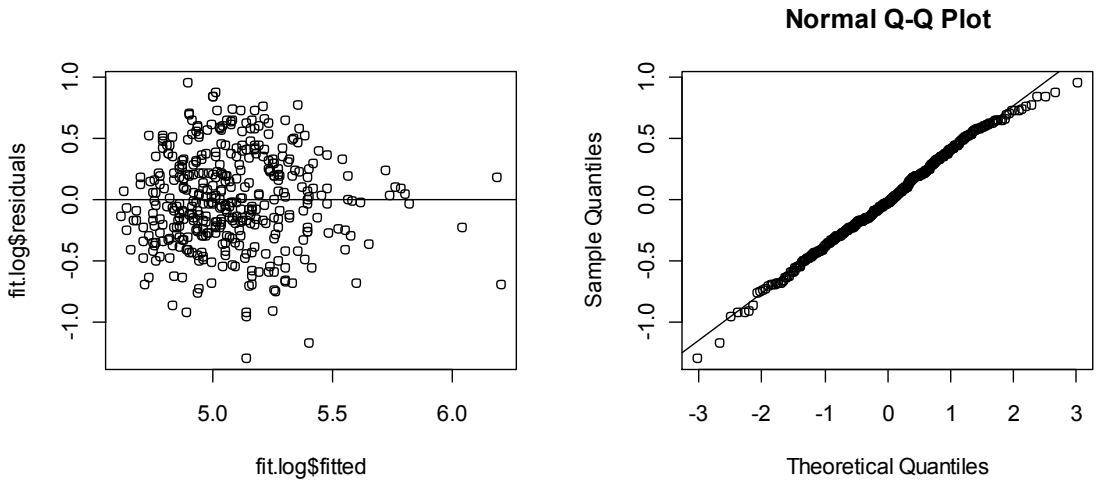
```
> # load the gee() package for robust standard errors
> library(gee)
>
> # identify outliers in scatterplot of triglycerides vs BMI
> plot(BMI,TG)
> bmi37 = which(BMI<=37)
>
> # excluding subjects with BMI > 37
> fit2 = lm(TG[bmi37] ~ BMI[bmi37])
> summary(fit2)

Call:
lm(formula = TG[bmi37] ~ BMI[bmi37])
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-169.07  -44.87  -13.22   39.45  232.05

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    -202.707     30.084  -6.738 5.68e-11 ***
BMI[bmi37]      15.199      1.199  12.677  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 68.01 on 396 degrees of freedom
Multiple R-Squared: 0.2887,     Adjusted R-squared: 0.2869
F-statistic: 160.7 on 1 and 396 DF,  p-value: < 2.2e-16

>
> # analyze residuals from the regression analysis of triglycerides and BMI
> plot(fit2$fitted, fit2$residuals)
> abline(0,0)
> qqnorm(fit2$residuals)
> qqline(fit2$residuals)
```



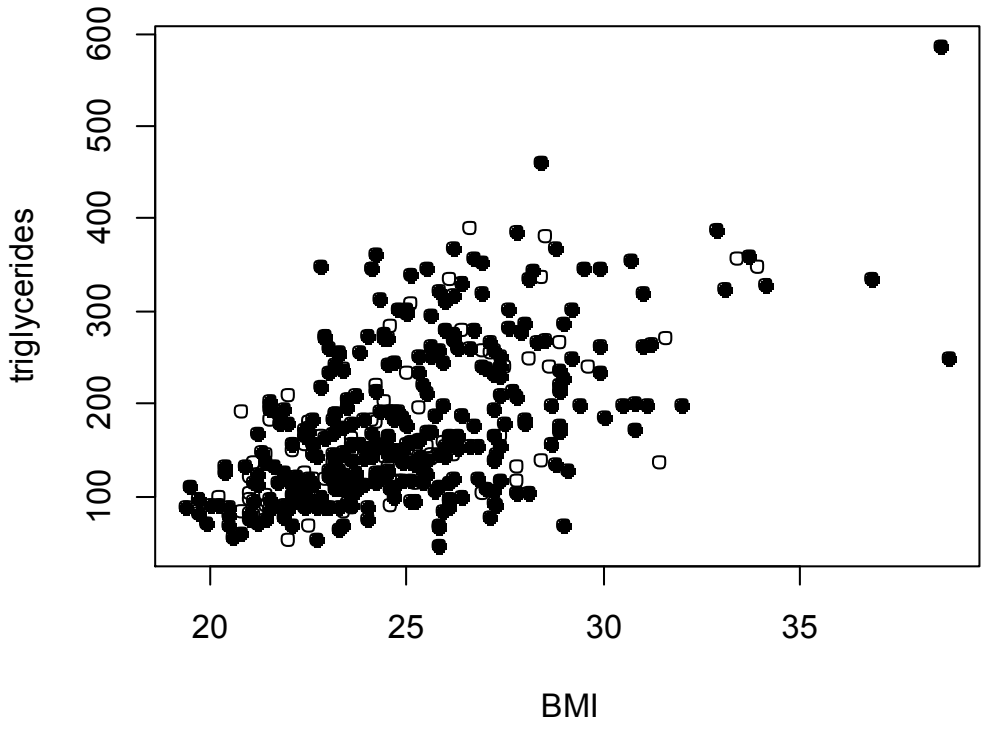

```
> # fit a linear regression model with robust standard errors
> fit.gee = gee(TG ~ BMI, id = seq(1,length(TG)))
[1] "Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27"
[1] "running glm to get initial regression estimate"
[1] -208.50096   15.43748
> summary(fit.gee)

GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
 Link:                      Identity
 Variance to Mean Relation: Gaussian
 Correlation Structure:     Independent

Call:
gee(formula = TG ~ BMI, id = seq(1, length(TG)))

Summary of Residuals:
       Min         1Q     Median         3Q        Max
-170.18608  -45.09554  -12.88618   39.60133  231.07641


Coefficients:
             Estimate Naive S.E.   Naive z Robust S.E.  Robust z
(Intercept) -208.50096  28.946250 -7.203039   32.021396 -6.511301
```

```
BMI           15.43748   1.149603 13.428538    1.322308 11.674646

Estimated Scale Parameter:  4750.958
Number of Iterations:  1

Working Correlation
     [,1]
[1,]    1
# calculate p-values for robust regression
> z = abs(fit.gee$coef/sqrt(diag(fit.gee$robust)))
> 2*(1-pnorm(z))
 (Intercept)          BMI
7.450263e-11 0.000000e+00
>
> # fit a regression model for log transformed triglycerides and BMI
> fit.log = lm(log(TG) ~ BMI)
> summary(fit.log)

Call:
lm(formula = log(TG) ~ BMI)

Residuals:
     Min       1Q   Median       3Q      Max
-1.29019 -0.25303 -0.01692  0.26530  0.95800

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.023584   0.162175   18.64   <2e-16 ***
BMI         0.082045   0.006441   12.74   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3862 on 398 degrees of freedom
Multiple R-Squared: 0.2896,     Adjusted R-squared: 0.2878
F-statistic: 162.3 on 1 and 398 DF,  p-value: < 2.2e-16

>
> # analyze residuals from the regression analysis of log transformed
> # triglycerides and BMI
> par(mfrow = c(1,2))
> plot(fit.log$fitted, fit.log$residuals)
> abline(0,0)
> qqnorm(fit.log$residuals)
> qqline(fit.log$residuals)
```




```
> # binary variable indicating presence of APOE4
> APOE4 = ifelse(APOE %in% c(3,5,6), 1, 0)
>
> # scatterplot with subjects stratified by APOE4
```

```
> par(mfrow = c(1,1))
> plot(BMI[APOE4 == 0], TG[APOE4 == 0], pch = 19, xlab = "BMI", ylab = "triglycerides")
> points(BMI[APOE4 == 1], TG[APOE4 == 1], pch = 1)
>
```




```
> # multiple linear regression of triglycerides on BMI, APOE4, and interaction
> fit3 = lm(TG ~ BMI + APOE4 + BMI*APOE4)
> summary(fit3)

Call:
lm(formula = TG ~ BMI + APOE4 + BMI * APOE4)

Residuals:

    Min      1Q  Median      3Q     Max
-170.04  -45.72  -13.03   38.88  231.12

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    -204.0193    32.4558  -6.286  8.6e-10 ***
BMI             15.2780     1.2857  11.883  < 2e-16 ***
APOE4           -20.9439    72.6801  -0.288    0.773
BMI:APOE4         0.7464     2.9088   0.257    0.798
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 69.09 on 396 degrees of freedom
Multiple R-Squared: 0.3121,   Adjusted R-squared: 0.3068
F-statistic: 59.88 on 3 and 396 DF,  p-value: < 2.2e-16
```

**ANOVA Lab 1**

---

The goal of this lab is to answer the following scientific questions using the cholesterol dataset:

- Is rs4775401 associated with cholesterol levels?
- Is APOE associated with cholesterol levels?

1. Load packages `multcomp` and `gee`

2. Perform a descriptive analysis to investigate the scientific questions of interest using numeric and graphical methods.

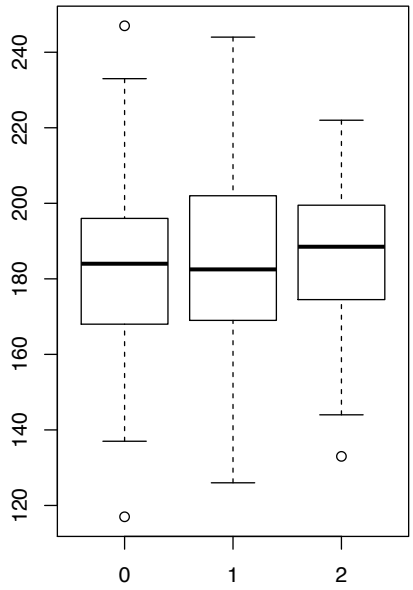
3. Compare the mean cholesterol levels between genotype groups defined by rs4775401.
    a. Perform the one-way ANOVA using the regression approach.
    b. Compare the above results with those obtained when
        i. allowing for unequal variances
        ii. using robust standard errors
        iii. using a nonparametric test
    c. Is there evidence that mean cholesterol levels between genotype groups are different? If so, perform all pairwise multiple comparisons using Bonferroni's adjustment. Try out different adjustment methods too.
    d. Interpret your results

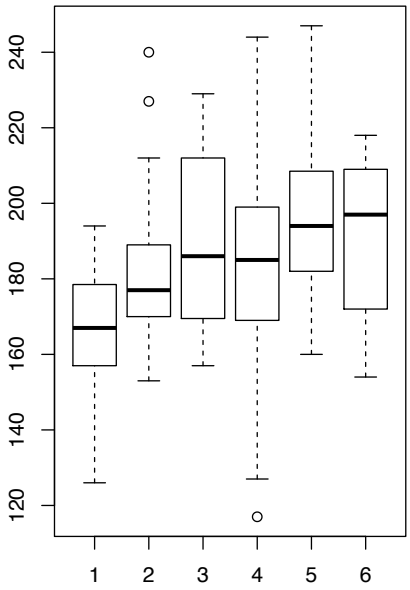
4. Repeat the steps described in problem 4 to compare the mean cholesterol levels between genotype groups defined by APOE.

---

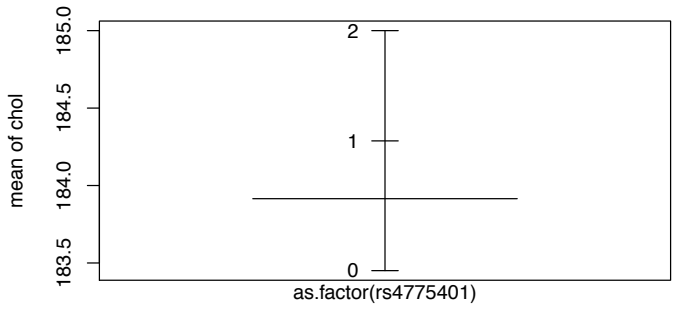
**R Commands & Output:**

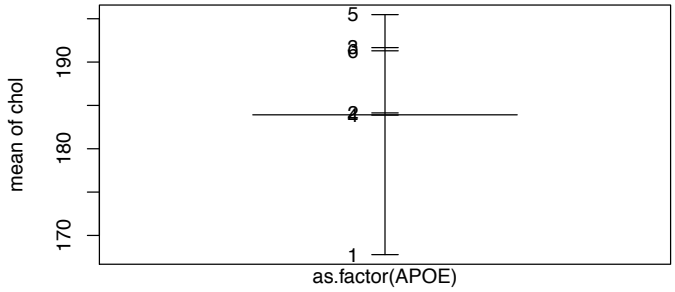
```
> library(multcomp)
> library(gee)
>
> ## read data set ------------------------------------------------------------
> cholesterol = read.table("http://faculty.washington.edu/rhubb/sisg/SISG-Data-
cholesterol.txt", header=T)
> attach(cholesterol)
>
> ## Exploratory data analysis ------------------------------------------------
> ## graphical display: boxplot
> par(mfrow = c(1,2))
> boxplot(chol ~ as.factor(rs4775401))
> boxplot(chol ~ as.factor(APOE))
```




```
> ## alternative graphical display: graph of means
> par(mfrow = c(2,1))
> plot.design(chol ~ as.factor(rs4775401))
> plot.design(chol ~ as.factor(APOE))
```




```
> ## numeric descriptives
> tapply(chol, as.factor(rs4775401), mean)
       0        1        2
183.4505 184.2882 185.0000
> tapply(chol, as.factor(rs4775401), sd)
       0        1        2
20.70619 23.85693 21.70851
>
> tapply(chol, as.factor(APOE), mean)
       1        2        3        4        5        6
167.7843 184.1364 191.6667 183.8826 195.4727 191.3000
> tapply(chol, as.factor(APOE), sd)
       1        2        3        4        5        6
15.70008 22.01146 26.20705 22.08516 17.22601 23.56575
>
> ## Inferential data analysis ------------------------------------------------
> fit1 = lm(chol ~ as.factor(rs4775401))
> summary(fit1)

Call:
lm(formula = chol ~ as.factor(rs4775401))

Residuals:
     Min       1Q   Median       3Q      Max
-66.4505 -15.4505  -0.2882  15.5495  63.5495

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          183.4505     1.5597 117.618   <2e-16 ***
as.factor(rs4775401)1   0.8377     2.3072   0.363    0.717
as.factor(rs4775401)2   1.5495     4.4702   0.347    0.729
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 22.17 on 397 degrees of freedom
Multiple R-squared: 0.0005135,  Adjusted R-squared: -0.004522
F-statistic: 0.102 on 2 and 397 DF,  p-value: 0.903

> anova(fit1)
Analysis of Variance Table

Response: chol
                      Df Sum Sq Mean Sq F value Pr(>F)
as.factor(rs4775401)   2    100   50.11   0.102 0.9031
Residuals            397 195089     491
>
> summary(fit2)

Call:
lm(formula = chol ~ as.factor(APOE))

Residuals:
   Min     1Q Median     3Q    Max
-66.88 -13.88  -0.46  15.12  60.12

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       167.784      2.935  57.162  < 2e-16 ***
as.factor(APOE)2   16.352      5.347   3.058 0.002378 **
as.factor(APOE)3   23.882      6.157   3.879 0.000123 ***
as.factor(APOE)4   16.098      3.224   4.993 8.94e-07 ***
as.factor(APOE)5   27.688      4.075   6.795 4.02e-11 ***
as.factor(APOE)6   23.516      7.250   3.244 0.001280 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.96 on 394 degrees of freedom
Multiple R-squared:  0.113,   Adjusted R-squared:  0.1018
F-statistic: 10.04 on 5 and 394 DF,  p-value: 4.616e-09

> anova(fit2)
Analysis of Variance Table

Response: chol
                Df Sum Sq Mean Sq F value    Pr(>F)
as.factor(APOE)   5  22065  4413.0  10.043 4.616e-09 ***
Residuals       394 173124   439.4
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

>
> ## all pairwise comparisons with different methods for adjustment
> M2 = contrMat(table(APOE), type="Tukey")
> fit3 = lm(chol ~ -1 + as.factor(APOE))
> mc2 = glht(fit3, linfct =M2)
> summary(mc2, test=adjusted("none"))

         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = chol ~ -1 + as.factor(APOE))

Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
2 - 1 == 0  16.3520     5.3468   3.058 0.002378 **
3 - 1 == 0  23.8824     6.1570   3.879 0.000123 ***
4 - 1 == 0  16.0983     3.2241   4.993 8.94e-07 ***
5 - 1 == 0  27.6884     4.0749   6.795 4.02e-11 ***
6 - 1 == 0  23.5157     7.2495   3.244 0.001280 **
3 - 2 == 0   7.5303     7.0190   1.073 0.283996
```

```
4 - 2 == 0  -0.2538     4.6639  -0.054 0.956634
5 - 2 == 0  11.3364     5.2879   2.144 0.032658 *
6 - 2 == 0   7.1636     7.9946   0.896 0.370765
4 - 3 == 0  -7.7841     5.5743  -1.396 0.163370
5 - 3 == 0   3.8061     6.1059   0.623 0.533423
6 - 3 == 0  -0.3667     8.5577  -0.043 0.965846
5 - 4 == 0  11.5901     3.1254   3.708 0.000239 ***
6 - 4 == 0   7.4174     6.7616   1.097 0.273315
6 - 5 == 0  -4.1727     7.2062  -0.579 0.562888
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- none method)

> summary(mc2, test=adjusted("bonferroni"))

         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = chol ~ -1 + as.factor(APOE))

Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
2 - 1 == 0  16.3520     5.3468   3.058  0.03567 *
3 - 1 == 0  23.8824     6.1570   3.879  0.00184 **
4 - 1 == 0  16.0983     3.2241   4.993 1.34e-05 ***
5 - 1 == 0  27.6884     4.0749   6.795 6.03e-10 ***
6 - 1 == 0  23.5157     7.2495   3.244  0.01920 *
3 - 2 == 0   7.5303     7.0190   1.073  1.00000
4 - 2 == 0  -0.2538     4.6639  -0.054  1.00000
5 - 2 == 0  11.3364     5.2879   2.144  0.48987
6 - 2 == 0   7.1636     7.9946   0.896  1.00000
4 - 3 == 0  -7.7841     5.5743  -1.396  1.00000
5 - 3 == 0   3.8061     6.1059   0.623  1.00000
6 - 3 == 0  -0.3667     8.5577  -0.043  1.00000
5 - 4 == 0  11.5901     3.1254   3.708  0.00358 **
6 - 4 == 0   7.4174     6.7616   1.097  1.00000
6 - 5 == 0  -4.1727     7.2062  -0.579  1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- bonferroni method)

> summary(mc2, test=adjusted("holm"))

         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = chol ~ -1 + as.factor(APOE))

Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
2 - 1 == 0  16.3520     5.3468   3.058  0.02378 *
3 - 1 == 0  23.8824     6.1570   3.879  0.00160 **
4 - 1 == 0  16.0983     3.2241   4.993 1.25e-05 ***
5 - 1 == 0  27.6884     4.0749   6.795 6.03e-10 ***
6 - 1 == 0  23.5157     7.2495   3.244  0.01408 *
3 - 2 == 0   7.5303     7.0190   1.073  1.00000
4 - 2 == 0  -0.2538     4.6639  -0.054  1.00000
5 - 2 == 0  11.3364     5.2879   2.144  0.29392
6 - 2 == 0   7.1636     7.9946   0.896  1.00000
4 - 3 == 0  -7.7841     5.5743  -1.396  1.00000
5 - 3 == 0   3.8061     6.1059   0.623  1.00000
6 - 3 == 0  -0.3667     8.5577  -0.043  1.00000
5 - 4 == 0  11.5901     3.1254   3.708  0.00286 **
6 - 4 == 0   7.4174     6.7616   1.097  1.00000
6 - 5 == 0  -4.1727     7.2062  -0.579  1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
(Adjusted p values reported -- holm method)

> summary(mc2, test=adjusted("hochberg"))

         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = chol ~ -1 + as.factor(APOE))

Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
2 - 1 == 0  16.3520     5.3468   3.058  0.02378 *
3 - 1 == 0  23.8824     6.1570   3.879  0.00160 **
4 - 1 == 0  16.0983     3.2241   4.993 1.25e-05 ***
5 - 1 == 0  27.6884     4.0749   6.795 6.03e-10 ***
6 - 1 == 0  23.5157     7.2495   3.244  0.01408 *
3 - 2 == 0   7.5303     7.0190   1.073  0.96585
4 - 2 == 0  -0.2538     4.6639  -0.054  0.96585
5 - 2 == 0  11.3364     5.2879   2.144  0.29392
6 - 2 == 0   7.1636     7.9946   0.896  0.96585
4 - 3 == 0  -7.7841     5.5743  -1.396  0.96585
5 - 3 == 0   3.8061     6.1059   0.623  0.96585
6 - 3 == 0  -0.3667     8.5577  -0.043  0.96585
5 - 4 == 0  11.5901     3.1254   3.708  0.00286 **
6 - 4 == 0   7.4174     6.7616   1.097  0.96585
6 - 5 == 0  -4.1727     7.2062  -0.579  0.96585
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- hochberg method)

> summary(mc2, test=adjusted("hommel"))

         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = chol ~ -1 + as.factor(APOE))

Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
2 - 1 == 0  16.3520     5.3468   3.058  0.02378 *
3 - 1 == 0  23.8824     6.1570   3.879  0.00155 **
4 - 1 == 0  16.0983     3.2241   4.993 1.25e-05 ***
5 - 1 == 0  27.6884     4.0749   6.795 6.03e-10 ***
6 - 1 == 0  23.5157     7.2495   3.244  0.01308 *
3 - 2 == 0   7.5303     7.0190   1.073  0.96585
4 - 2 == 0  -0.2538     4.6639  -0.054  0.96585
5 - 2 == 0  11.3364     5.2879   2.144  0.29392
6 - 2 == 0   7.1636     7.9946   0.896  0.96585
4 - 3 == 0  -7.7841     5.5743  -1.396  0.84433
5 - 3 == 0   3.8061     6.1059   0.623  0.96585
6 - 3 == 0  -0.3667     8.5577  -0.043  0.96585
5 - 4 == 0  11.5901     3.1254   3.708  0.00286 **
6 - 4 == 0   7.4174     6.7616   1.097  0.96585
6 - 5 == 0  -4.1727     7.2062  -0.579  0.96585
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- hommel method)

> summary(mc2, test=adjusted("BH"))

         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = chol ~ -1 + as.factor(APOE))
```

```
Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
2 - 1 == 0  16.3520     5.3468   3.058 0.005945 **
3 - 1 == 0  23.8824     6.1570   3.879 0.000615 ***
4 - 1 == 0  16.0983     3.2241   4.993 6.70e-06 ***
5 - 1 == 0  27.6884     4.0749   6.795 6.03e-10 ***
6 - 1 == 0  23.5157     7.2495   3.244 0.003841 **
3 - 2 == 0   7.5303     7.0190   1.073 0.425994
4 - 2 == 0  -0.2538     4.6639  -0.054 0.965846
5 - 2 == 0  11.3364     5.2879   2.144 0.069981 .
6 - 2 == 0   7.1636     7.9946   0.896 0.505589
4 - 3 == 0  -7.7841     5.5743  -1.396 0.306319
5 - 3 == 0   3.8061     6.1059   0.623 0.649486
6 - 3 == 0  -0.3667     8.5577  -0.043 0.965846
5 - 4 == 0  11.5901     3.1254   3.708 0.000894 ***
6 - 4 == 0   7.4174     6.7616   1.097 0.425994
6 - 5 == 0  -4.1727     7.2062  -0.579 0.649486
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- BH method)

> summary(mc2, test=adjusted("BY"))

         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = chol ~ -1 + as.factor(APOE))

Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
2 - 1 == 0  16.3520     5.3468   3.058  0.01973 *
3 - 1 == 0  23.8824     6.1570   3.879  0.00204 **
4 - 1 == 0  16.0983     3.2241   4.993 2.22e-05 ***
5 - 1 == 0  27.6884     4.0749   6.795 2.00e-09 ***
6 - 1 == 0  23.5157     7.2495   3.244  0.01274 *
3 - 2 == 0   7.5303     7.0190   1.073  1.00000
4 - 2 == 0  -0.2538     4.6639  -0.054  1.00000
5 - 2 == 0  11.3364     5.2879   2.144  0.23221
6 - 2 == 0   7.1636     7.9946   0.896  1.00000
4 - 3 == 0  -7.7841     5.5743  -1.396  1.00000
5 - 3 == 0   3.8061     6.1059   0.623  1.00000
6 - 3 == 0  -0.3667     8.5577  -0.043  1.00000
5 - 4 == 0  11.5901     3.1254   3.708  0.00297 **
6 - 4 == 0   7.4174     6.7616   1.097  1.00000
6 - 5 == 0  -4.1727     7.2062  -0.579  1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- BY method)

> summary(mc2, test=adjusted("fdr"))

         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = chol ~ -1 + as.factor(APOE))

Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
2 - 1 == 0  16.3520     5.3468   3.058 0.005945 **
3 - 1 == 0  23.8824     6.1570   3.879 0.000615 ***
4 - 1 == 0  16.0983     3.2241   4.993 6.70e-06 ***
5 - 1 == 0  27.6884     4.0749   6.795 6.03e-10 ***
6 - 1 == 0  23.5157     7.2495   3.244 0.003841 **
3 - 2 == 0   7.5303     7.0190   1.073 0.425994
4 - 2 == 0  -0.2538     4.6639  -0.054 0.965846
5 - 2 == 0  11.3364     5.2879   2.144 0.069981 .
6 - 2 == 0   7.1636     7.9946   0.896 0.505589
```

```
4 - 3 == 0  -7.7841     5.5743  -1.396 0.306319
5 - 3 == 0   3.8061     6.1059   0.623 0.649486
6 - 3 == 0  -0.3667     8.5577  -0.043 0.965846
5 - 4 == 0  11.5901     3.1254   3.708 0.000894 ***
6 - 4 == 0   7.4174     6.7616   1.097 0.425994
6 - 5 == 0  -4.1727     7.2062  -0.579 0.649486
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- fdr method)
>
> ## One-way (not assuming equal variances)
> oneway.test(chol ~ as.factor(rs4775401))

        One-way analysis of means (not assuming equal variances)

data:  chol and as.factor(rs4775401)
F = 0.1046, num df = 2.000, denom df = 75.608, p-value = 0.9008

>
> oneway.test(chol ~ as.factor(APOE))

        One-way analysis of means (not assuming equal variances)

data:  chol and as.factor(APOE)
F = 15.167, num df = 5.00, denom df = 47.75, p-value = 6.524e-09
>
> ## Using robust standard errors
> summary(gee(chol ~ as.factor(rs4775401), id=seq(1,length(chol))))
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
          (Intercept) as.factor(rs4775401)1 as.factor(rs4775401)2
          183.4504950             0.8377402             1.5495050

 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
 Link:                      Identity
 Variance to Mean Relation: Gaussian
 Correlation Structure:     Independent

Call:
gee(formula = chol ~ as.factor(rs4775401), id = seq(1, length(chol)))

Summary of Residuals:
        Min          1Q      Median          3Q         Max
-66.4504950 -15.4504950  -0.2882353  15.5495050  63.5495050


Coefficients:
                      Estimate Naive S.E.     Naive z Robust S.E.
(Intercept)          183.4504950   1.559715 117.6179395    1.453272
as.factor(rs4775401)1   0.8377402   2.307238   0.3630923    2.332437
as.factor(rs4775401)2   1.5495050   4.470234   0.3466273    4.282708
                       Robust z
(Intercept)          126.2327489
as.factor(rs4775401)1   0.3591694
as.factor(rs4775401)2   0.3618049

Estimated Scale Parameter:  491.4078
Number of Iterations:  1

Working Correlation
     [,1]
[1,]    1
>
> summary(gee(chol ~ as.factor(APOE), id=seq(1,length(chol))))
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
     (Intercept) as.factor(APOE)2 as.factor(APOE)3 as.factor(APOE)4 as.factor(APOE)5
as.factor(APOE)6
```

```
       167.78431         16.35205         23.88235         16.09828         27.68841
23.51569

 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
 Link:                      Identity
 Variance to Mean Relation: Gaussian
 Correlation Structure:     Independent

Call:
gee(formula = chol ~ as.factor(APOE), id = seq(1, length(chol)))

Summary of Residuals:
        Min          1Q      Median          3Q         Max
-66.8825911 -13.8825911  -0.4603387  15.1174089  60.1174089


Coefficients:
                   Estimate Naive S.E.   Naive z Robust S.E.  Robust z
(Intercept)      167.78431   2.935252 57.161810    2.176791 77.078744
as.factor(APOE)2  16.35205   5.346819  3.058276    5.075461  3.221786
as.factor(APOE)3  23.88235   6.157036  3.878871    6.890082  3.466193
as.factor(APOE)4  16.09828   3.224074  4.993147    2.589428  6.216924
as.factor(APOE)5  27.68841   4.074900  6.794869    3.167888  8.740339
as.factor(APOE)6  23.51569   7.249538  3.243750    7.397258  3.178974

Estimated Scale Parameter:  439.4009
Number of Iterations:  1

Working Correlation
     [,1]
[1,]    1
> ## non-parametric ANOVA
> kruskal.test(chol ~ as.factor(rs4775401))

        Kruskal-Wallis rank sum test

data:  chol by as.factor(rs4775401)
Kruskal-Wallis chi-squared = 0.5761, df = 2, p-value = 0.7497


> kruskal.test(chol ~ as.factor(APOE))

        Kruskal-Wallis rank sum test

data:  chol by as.factor(APOE)
Kruskal-Wallis chi-squared = 48.227, df = 5, p-value = 3.193e-09
```

## ANOVA Lab 2

---

The goal of this lab is to answer the following scientific questions using the cholesterol dataset.

- Are rs174548 and APOE associated with cholesterol levels?
- Does the effect of APOE on cholesterol levels depend on rs174548?

1. Obtain a cross-tabulation of the groups defined by rs174548 and APOE.

2. Perform a descriptive analysis to investigate the scientific questions of interest using numeric and graphical methods.

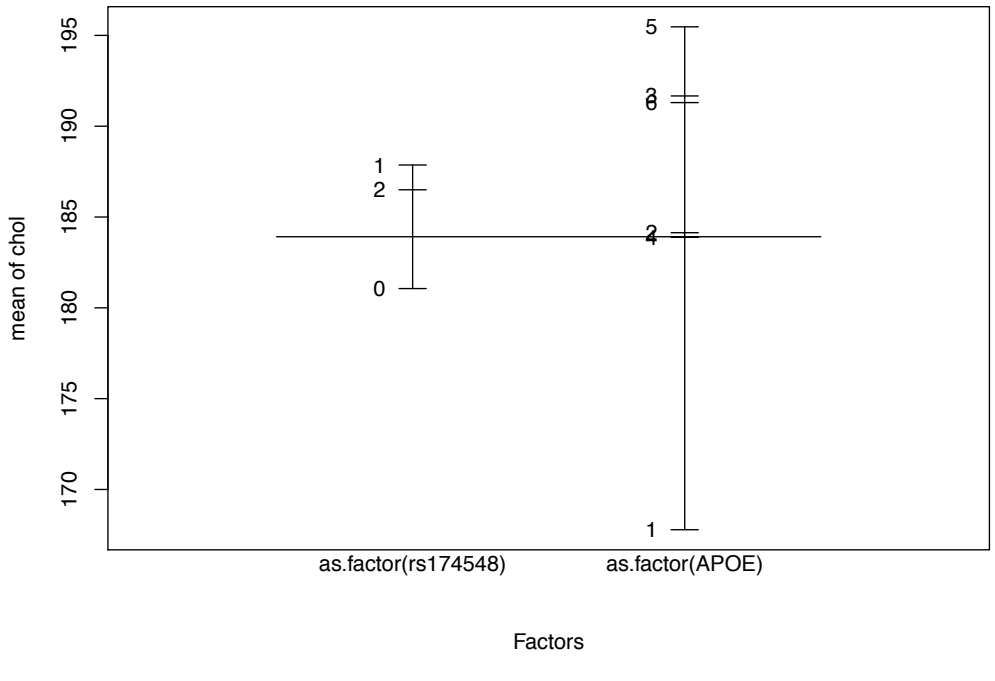
3. Fit a two-way ANOVA model with an interaction between rs174548 and APOE. Test the interaction. What do you conclude?

4. Fit a two-way ANOVA model without the interaction between rs174548 and APOE. Test the main effects of rs174548 and APOE. What do you conclude?

---

**R Commands & Output:**

```
> ## Two-way ANOVA ------------------------------------------------------------
> ## exploratory data analysis
> table(rs174548, APOE)
        APOE
rs174548   1   2   3   4   5   6
       0  33  10   8 136  34   6
       1  17   9   7  90  20   4
       2   1   3   0  21   1   0
> tapply(chol, list(as.factor(rs174548), as.factor(APOE)), mean)
         1        2        3        4        5        6
0 168.0909 179.7000 198.1250 180.4706 192.4706 180.6667
1 167.7059 187.2222 184.2857 187.9889 202.1000 207.2500
2 159.0000 189.6667       NA 188.3810 165.0000       NA
> tapply(chol, list(as.factor(rs174548), as.factor(APOE)), sd)
         1        2        3        4        5        6
0 17.39318 19.72618 26.15032 21.06531 15.66164 23.04488
1 12.65783 26.54608 26.18342 23.93460 17.49556 14.68276
2       NA 18.17507       NA 16.68975       NA       NA
>
> plot.design(chol ~ as.factor(rs174548) + as.factor(APOE))
```




```
> ## model with interaction
> fit1 = lm(chol ~ as.factor(rs174548)*as.factor(APOE))
> summary(fit1)

Call:
lm(formula = chol ~ as.factor(rs174548) * as.factor(APOE))

Residuals:
   Min     1Q Median     3Q    Max
-63.47 -12.53  -0.24  13.91  56.01

Coefficients: (2 not defined because of singularities)
                                     Estimate Std. Error t value Pr(>|t|)
(Intercept)                           168.091      3.604  46.637  < 2e-16 ***
as.factor(rs174548)1                   -0.385      6.181  -0.062 0.950365
as.factor(rs174548)2                   -9.091     21.016  -0.433 0.665574
as.factor(APOE)2                       11.609      7.474   1.553 0.121183
as.factor(APOE)3                       30.034      8.159   3.681 0.000266 ***
as.factor(APOE)4                       12.380      4.018   3.081 0.002210 **
as.factor(APOE)5                       24.380      5.060   4.819 2.09e-06 ***
as.factor(APOE)6                       12.576      9.189   1.369 0.171941
as.factor(rs174548)1:as.factor(APOE)2   7.907     11.345   0.697 0.486239
as.factor(rs174548)2:as.factor(APOE)2  19.058     25.049   0.761 0.447237
as.factor(rs174548)1:as.factor(APOE)3 -13.454     12.371  -1.088 0.277463
as.factor(rs174548)2:as.factor(APOE)3      NA         NA      NA       NA
as.factor(rs174548)1:as.factor(APOE)4   7.903      6.791   1.164 0.245260
as.factor(rs174548)2:as.factor(APOE)4  17.001     21.570   0.788 0.431065
as.factor(rs174548)1:as.factor(APOE)5  10.014      8.500   1.178 0.239464
as.factor(rs174548)2:as.factor(APOE)5 -18.380     29.715  -0.619 0.536594
as.factor(rs174548)1:as.factor(APOE)6  26.968     14.725   1.831 0.067809 .
as.factor(rs174548)2:as.factor(APOE)6      NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.7 on 384 degrees of freedom
Multiple R-squared:  0.1566,  Adjusted R-squared:  0.1237
F-statistic: 4.754 on 15 and 384 DF,  p-value: 1.865e-08
```

```
>
> ## model without interaction
> fit2 = lm(chol ~ as.factor(rs174548) + as.factor(APOE))
> summary(fit2)

Call:
lm(formula = chol ~ as.factor(rs174548) + as.factor(APOE))

Residuals:
    Min      1Q  Median      3Q     Max
-64.070 -13.070  -0.465  14.519  56.519

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           165.538      3.006  55.060  < 2e-16 ***
as.factor(rs174548)1    6.410      2.204   2.908 0.003844 **
as.factor(rs174548)2    5.604      4.354   1.287 0.198831
as.factor(APOE)2       15.212      5.330   2.854 0.004548 **
as.factor(APOE)3       23.138      6.110   3.787 0.000177 ***
as.factor(APOE)4       15.533      3.211   4.838 1.89e-06 ***
as.factor(APOE)5       27.502      4.040   6.808 3.74e-11 ***
as.factor(APOE)6       23.198      7.188   3.227 0.001354 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.78 on 392 degrees of freedom
Multiple R-squared:  0.1329,  Adjusted R-squared:  0.1174
F-statistic:  8.58 on 7 and 392 DF,  p-value: 8.485e-10

## compare models with and without interaction
> anova(fit2,fit1)
Analysis of Variance Table

Model 1: chol ~ as.factor(rs174548) + as.factor(APOE)
Model 2: chol ~ as.factor(rs174548) * as.factor(APOE)
  Res.Df    RSS Df Sum of Sq     F Pr(>F)
1    392 169256
2    384 164619  8    4636.7 1.352 0.2163
```

**ANOVA Lab 3**

---

The goal of this lab is to answer the following scientific questions using the cholesterol dataset.

- Controlling for age, is APOE associated with cholesterol levels?
- Does age modify the association between APOE and cholesterol levels?

1. Perform a descriptive analysis to investigate the scientific questions of interest using numeric and graphical methods.

2. Fit an ANCOVA model with an interaction between APOE and age. Test the interaction. What do you conclude?

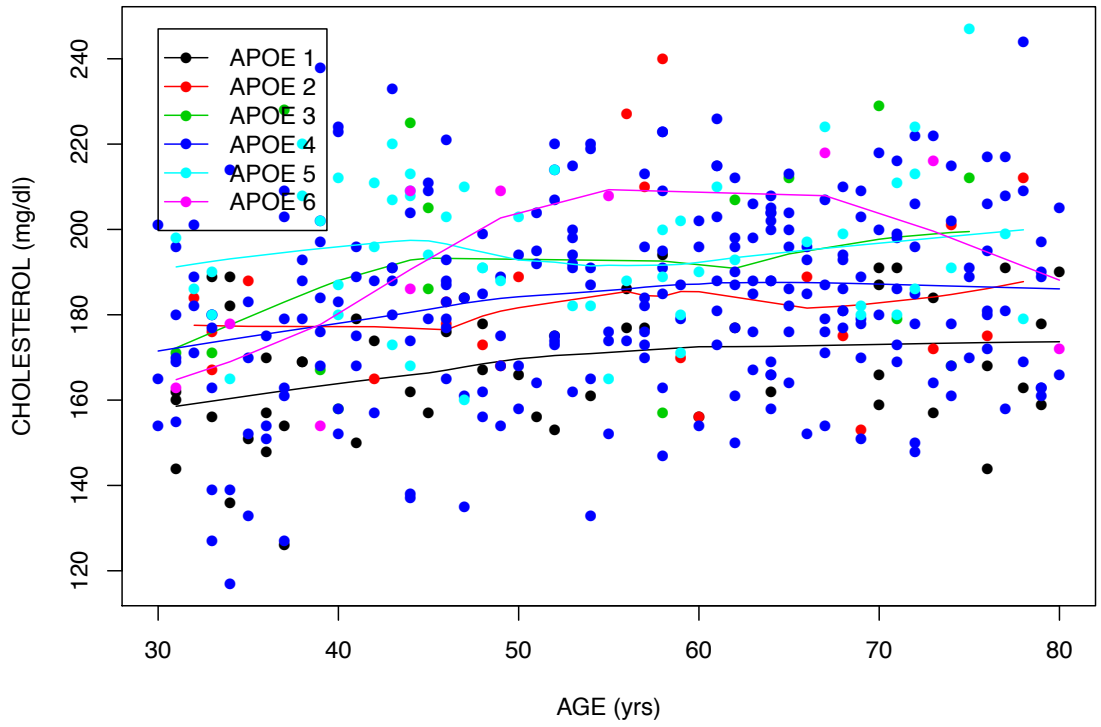
3. Fit an ANCOVA model without an interaction between APOE and age. Compare the results with the one-way ANOVA model that compares mean cholesterol levels among genotypes defined by APOE. What can you say about the role of age? [Is it an effect modifier? Or is it a confounder? Or is it a precision variable?]

---

**R Commands & Output:**

```
> by(cbind(chol,age), APOE, cor, method="pearson")
INDICES: 1
          chol       age
chol 1.0000000 0.3120186
age  0.3120186 1.0000000
------------------------------------------------------------------------------------------
------------------------------------------------------------
INDICES: 2
          chol       age
chol 1.0000000 0.1562559
age  0.1562559 1.0000000
------------------------------------------------------------------------------------------
------------------------------------------------------------
INDICES: 3
          chol       age
chol 1.0000000 0.2196231
age  0.2196231 1.0000000
------------------------------------------------------------------------------------------
------------------------------------------------------------
INDICES: 4
          chol       age
chol 1.0000000 0.2107872
age  0.2107872 1.0000000
------------------------------------------------------------------------------------------
------------------------------------------------------------
INDICES: 5
          chol       age
chol 1.0000000 0.1137494
age  0.1137494 1.0000000
------------------------------------------------------------------------------------------
------------------------------------------------------------
INDICES: 6
          chol       age
```

```
chol 1.0000000 0.4487348
age  0.4487348 1.0000000
> by(cbind(chol,age), APOE, cor, method="spearman")
INDICES: 1
          chol       age
chol 1.0000000 0.3139938
age  0.3139938 1.0000000
------------------------------------------------------------------------------------------------
------------------------------------------------------------
INDICES: 2
          chol       age
chol 1.0000000 0.1000565
age  0.1000565 1.0000000
------------------------------------------------------------------------------------------------
------------------------------------------------------------
INDICES: 3
          chol       age
chol 1.0000000 0.2184423
age  0.2184423 1.0000000
------------------------------------------------------------------------------------------------
------------------------------------------------------------
INDICES: 4
          chol       age
chol 1.0000000 0.1785631
age  0.1785631 1.0000000
------------------------------------------------------------------------------------------------
------------------------------------------------------------
INDICES: 5
           chol        age
chol 1.00000000 0.02929649
age  0.02929649 1.00000000
------------------------------------------------------------------------------------------------
------------------------------------------------------------
INDICES: 6
          chol       age
chol 1.0000000 0.5457317
age  0.5457317 1.0000000
>
>
> plot(age, chol, xlab="AGE (yrs)", ylab="CHOLESTEROL (mg/dl)", type="n")
> for (i in 1:6){
+  lines(lowess(age[APOE==i], chol[APOE==i]), col=i)
+  points(age[APOE==i], chol[APOE==i], col=i, pch=16)
+ }
> legend(min(age), max(chol), legend=paste("APOE", seq(1,6)), col=seq(1,6), pch=16,
lty=1)
>
```




```
> ## ANCOVA Model with an interaction
> fit1 = lm(chol ~ as.factor(APOE) * age)
> summary(fit1)

Call:
lm(formula = chol ~ as.factor(APOE) * age)

Residuals:
    Min      1Q  Median      3Q     Max
-59.605 -13.691   0.216  13.843  59.741

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)          151.94102    9.85418  15.419   <2e-16 ***
as.factor(APOE)2      19.34227   19.96554   0.969   0.3333
as.factor(APOE)3      20.57706   21.46917   0.958   0.3384
as.factor(APOE)4      13.41554   11.23143   1.194   0.2330
as.factor(APOE)5      35.78221   15.09145   2.371   0.0182 *
as.factor(APOE)6       6.80388   24.21700   0.281   0.7789
age                    0.30262    0.17998   1.681   0.0935 .
as.factor(APOE)2:age  -0.06855    0.35497  -0.193   0.8470
as.factor(APOE)3:age   0.06142    0.39200   0.157   0.8756
as.factor(APOE)4:age   0.02820    0.20274   0.139   0.8895
as.factor(APOE)5:age  -0.15636    0.27602  -0.566   0.5714
as.factor(APOE)6:age   0.32829    0.44749   0.734   0.4636
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.6 on 388 degrees of freedom
Multiple R-squared:  0.1564,  Adjusted R-squared:  0.1325
F-statistic: 6.538 on 11 and 388 DF,  p-value: 5.088e-10

>
> ## ANCOVA Model without an interaction
```

```
> fit2 = lm(chol ~ as.factor(APOE) + age)
> summary(fit2)

Call:
lm(formula = chol ~ as.factor(APOE) + age)

Residuals:
    Min      1Q  Median      3Q     Max
-60.066 -14.163   0.195  13.966  59.385

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      151.56305    4.72094  32.104  < 2e-16 ***
as.factor(APOE)2  15.56004    5.23358   2.973  0.00313 **
as.factor(APOE)3  23.80580    6.02298   3.952 9.17e-05 ***
as.factor(APOE)4  14.96826    3.16465   4.730 3.14e-06 ***
as.factor(APOE)5  27.49356    3.98641   6.897 2.13e-11 ***
as.factor(APOE)6  23.74898    7.09187   3.349  0.00089 ***
age                0.30984    0.07158   4.329 1.91e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.51 on 393 degrees of freedom
Multiple R-squared:  0.1534,  Adjusted R-squared:  0.1405
F-statistic: 11.87 on 6 and 393 DF,  p-value: 2.993e-12

>
> ## compare models with and without interaction
> anova(fit2, fit1)
Analysis of Variance Table

Model 1: chol ~ as.factor(APOE) + age
Model 2: chol ~ as.factor(APOE) * age
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    393 165245
2    388 164667  5    578.37 0.2726  0.928
>
> ## ONE-WAY ANOVA model
> fit3 = lm(chol ~ as.factor(APOE))
> summary(fit3)

Call:
lm(formula = chol ~ as.factor(APOE))

Residuals:
   Min     1Q Median     3Q    Max
-66.88 -13.88  -0.46  15.12  60.12

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       167.784      2.935  57.162  < 2e-16 ***
as.factor(APOE)2   16.352      5.347   3.058 0.002378 **
as.factor(APOE)3   23.882      6.157   3.879 0.000123 ***
as.factor(APOE)4   16.098      3.224   4.993 8.94e-07 ***
as.factor(APOE)5   27.688      4.075   6.795 4.02e-11 ***
as.factor(APOE)6   23.516      7.250   3.244 0.001280 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.96 on 394 degrees of freedom
Multiple R-squared:  0.113,   Adjusted R-squared:  0.1018
F-statistic: 10.04 on 5 and 394 DF,  p-value: 4.616e-09

> anova(fit3, fit2)
Analysis of Variance Table

Model 1: chol ~ as.factor(APOE)
Model 2: chol ~ as.factor(APOE) + age
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    394 173124
2    393 165245  1    7878.7 18.738 1.905e-05 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> ## mean cholesterol for different genotypes
> predict(fit3, new=data.frame(APOE=1))
       1
167.7843
> predict(fit3, new=data.frame(APOE=2))
       1
184.1364
> predict(fit3, new=data.frame(APOE=3))
       1
191.6667
> predict(fit3, new=data.frame(APOE=4))
       1
183.8826
> predict(fit3, new=data.frame(APOE=5))
       1
195.4727
> predict(fit3, new=data.frame(APOE=6))
    1
191.3
>
> ## mean cholesterol for different genotypes adjusted by age
> predict(fit2, new=data.frame(age=mean(age),APOE=1))
       1
168.5495
> predict(fit2, new=data.frame(age=mean(age),APOE=2))
       1
184.1095
> predict(fit2, new=data.frame(age=mean(age),APOE=3))
       1
192.3553
> predict(fit2, new=data.frame(age=mean(age),APOE=4))
       1
183.5177
> predict(fit2, new=data.frame(age=mean(age),APOE=5))
       1
196.0431
> predict(fit2, new=data.frame(age=mean(age),APOE=6))
       1
192.2985
```