

# Module 4: Regression Methods: Concepts and Applications

---

## Lab 1: Introduction to R and Simple Linear Regression

The goal of the lab sessions is to use the cholesterol data set to explore relationships among the variables. The objectives of this first lab are

- Become familiar with R and RStudio
- Begin to explore the cholesterol dataset
- Use graphical and descriptive methods to investigate the association between triglycerides and BMI
- Use simple linear regression to investigate the association between triglycerides and BMI

The cholesterol data set is available for download from the module Github repository and contains the following variables:

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<sup>2</sup>

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.

HTN: diagnosed hypertension: 0 = no, 1 = yes

chd: diagnosis of coronary heart disease: 0 = no, 1 = yes

You can download the data file and read it into R as follows:

```
cholesterol = read.csv("https://raw.githubusercontent.com/rhubb/SISG2018/master/data/SISG-D
ata-cholesterol.csv", header=T)
```

## Exercises

1. Provide descriptive statistics summarizing triglycerides and BMI.

```
summary(TG)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      47.0   114.8   156.5   177.4   234.0   586.0
```

```
summary(BMI)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      19.40   22.90   24.60   25.00   26.73   38.80
```

```
group = 1*(BMI > 25)
group=factor(group,levels=c(0,1), labels=c("<=25",">25"))
table(group)
```

```
## group
## <=25  >25
##   224   176
```

```
by(TG, group, mean)
```

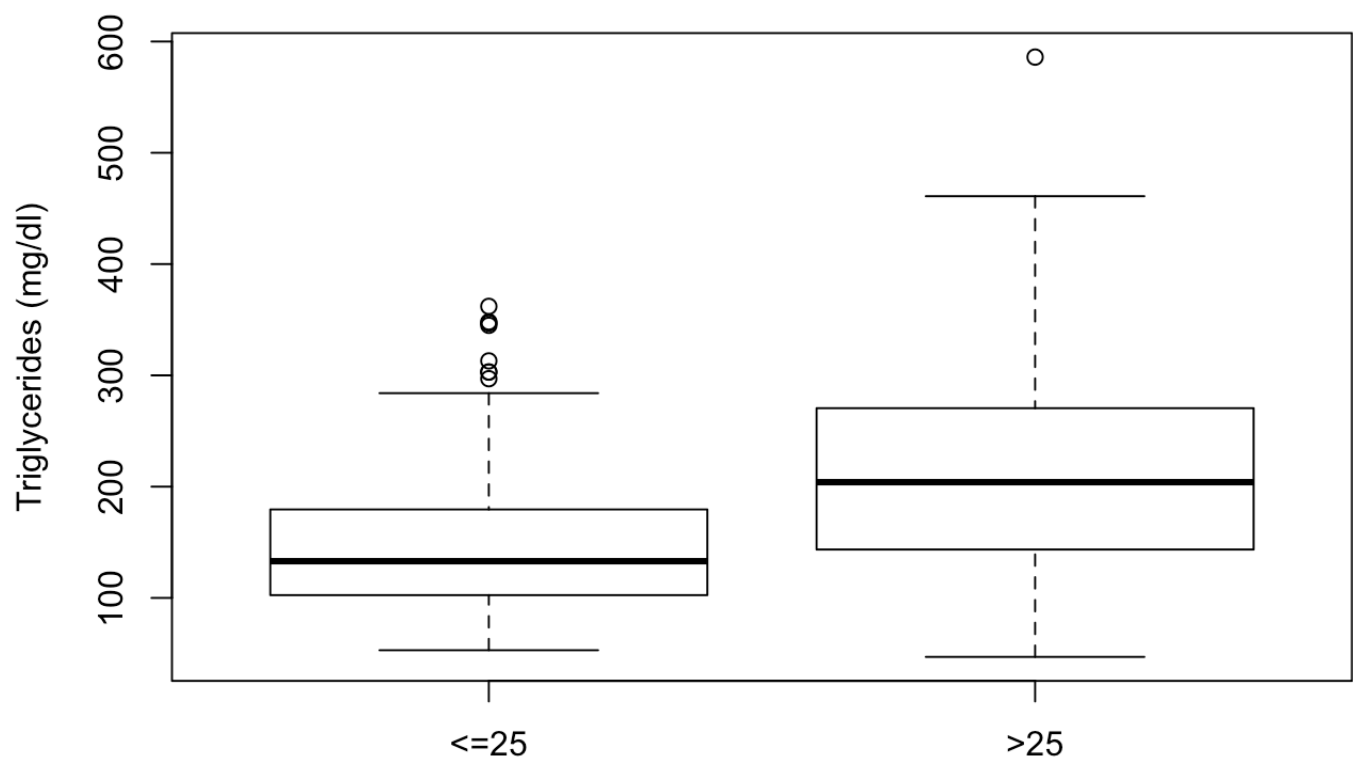
```
## group: <=25
## [1] 147.3839
## -----
## -----
## -----
## -----
## group: >25
## [1] 215.6932
```

```
by(TG, group, sd)
```

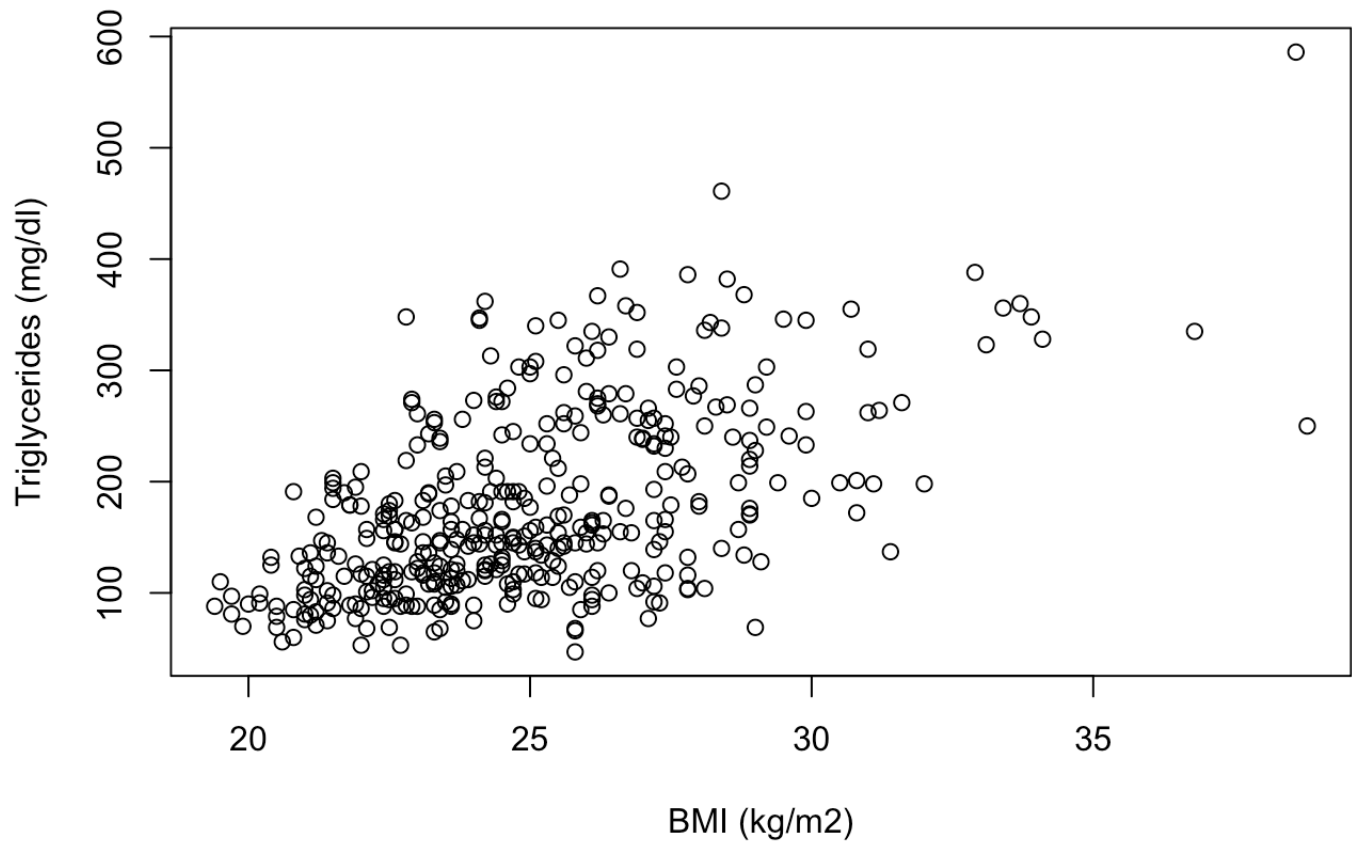
```
## group: <=25
## [1] 61.70787
## -----
## -----
## -----
## -----
## group: >25
## [1] 90.66584
```

2. Create plots illustrating the relationship between triglycerides and BMI. Based on your graphical summaries does there appear to be an association between triglycerides and BMI?

```
boxplot(TG~group,ylab="Triglycerides (mg/dl)")
```



```
plot(TG ~ BMI, xlab = "BMI (kg/m2)", ylab = "Triglycerides (mg/dl)")
```



3. Use linear regression to investigate the association between triglycerides and BMI. What do the linear regression model results tell us about the association? Make sure you can interpret the model coefficients and any hypothesis testing.

```
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
```

4. 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?

```
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
```

Once your group has completed the lab exercises, please submit your R script file to the class Github repository:

<https://github.com/rhubb/SISG2018/tree/master/submit> (<https://github.com/rhubb/SISG2018/tree/master/submit>)

Sign in using the class username and password. Then click upload files to save your R script file to the repository.