MA308PROJ_ANOVA

import dataset

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.0.3
```

```
heart <- read_csv("heart_cleveland_upload.csv")
```

```
##
## -- Column specification --------------------------------------
## cols(
##   age = col_double(),
##   sex = col_double(),
##   cp = col_double(),
##   trestbps = col_double(),
##   chol = col_double(),
##   fbs = col_double(),
##   restecg = col_double(),
##   thalach = col_double(),
##   exang = col_double(),
##   oldpeak = col_double(),
##   slope = col_double(),
##   ca = col_double(),
##   thal = col_double(),
##   condition = col_double()
## )
```

```
attach(heart)
```

I was wondering the sample size of each group differs a lot, so can ANOVA test still works?

Questions: 1.parallel factor we can see the condition's difference between the label; but for continuous number, we should see the CI and mean from the condition's aspect. 2.If we use One-way ANCOVA, does the importance of oldpeak and slope equal?

one-way ANOVA for CP

```
library(gplots)
```

```
## Warning: package 'gplots' was built under R version 4.0.3
```

```
##
## Attaching package: 'gplots'

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

```
library(HH)
```

```
## Warning: package 'HH' was built under R version 4.0.3

## Loading required package: lattice

## Loading required package: grid

## Loading required package: latticeExtra

## Loading required package: multcomp

## Warning: package 'multcomp' was built under R version 4.0.3

## Loading required package: mvtnorm

## Warning: package 'mvtnorm' was built under R version 4.0.3

## Loading required package: survival

##
## Attaching package: 'survival'

## The following object is masked _by_ '.GlobalEnv':
##
##     heart

## Loading required package: TH.data

## Warning: package 'TH.data' was built under R version 4.0.3

## Loading required package: MASS

## Warning: package 'MASS' was built under R version 4.0.3

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##     geyser
```

```
## Loading required package: gridExtra
```

```
##
## Attaching package: 'HH'
```

```
## The following object is masked from 'package:gplots':
##
##     residplot
```

```
library(multcomp)
library(car)
```

```
## Warning: package 'car' was built under R version 4.0.3
```

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

```
## Warning: package 'carData' was built under R version 4.0.3
```

```
##
## Attaching package: 'car'
```

```
## The following objects are masked from 'package:HH':
##
##     logit, vif
```

```
#table(cp)
aggregate(condition, by=list(cp),FUN=mean)
```

```
##   Group.1         x
## 1       0 0.3043478
## 2       1 0.1836735
## 3       2 0.2168675
## 4       3 0.7253521
```

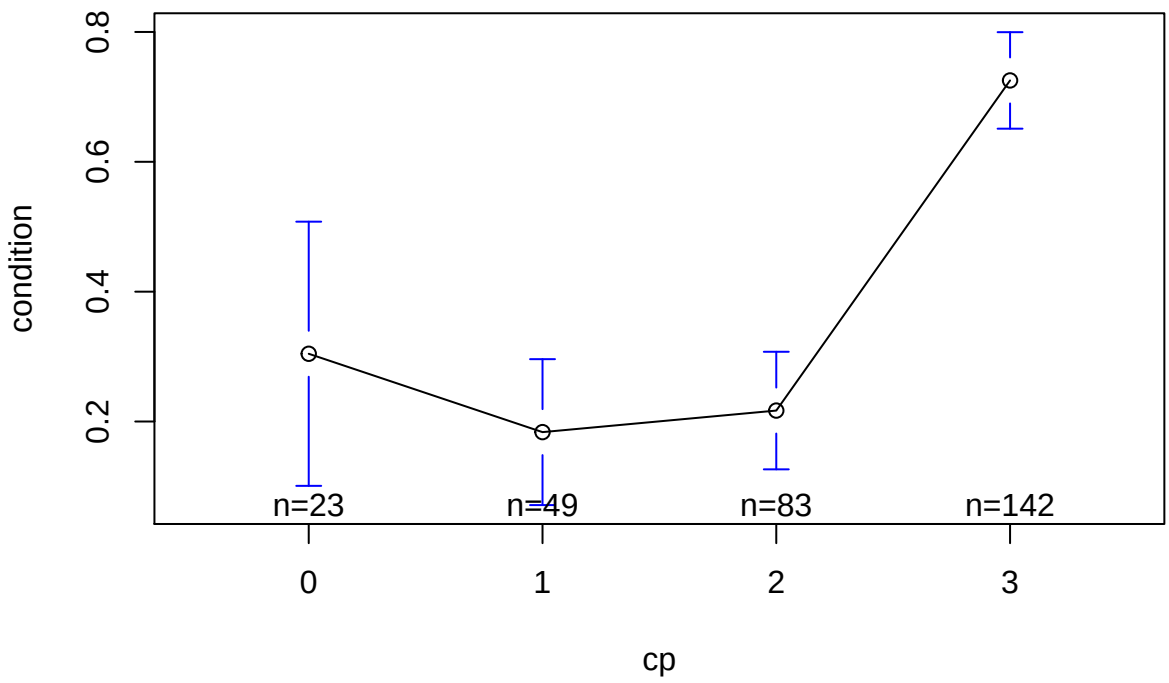
```
aggregate(condition, by=list(cp),FUN=sd)
```

```
##   Group.1         x
## 1       0 0.4704720
## 2       1 0.3912304
## 3       2 0.4146169
## 4       3 0.4479166
```

```
fitcp <- aov(condition ~ cp)
summary(fitcp)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## cp            1  12.34  12.343   59.24 2.11e-13 ***
## Residuals   295  61.46   0.208
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plotmeans(condition ~ cp)
```




```
#TukeyHSD(fitcp)
bartlett.test(condition ~ cp)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  condition by cp
## Bartlett's K-squared = 1.8534, df = 3, p-value = 0.6034
```

one-way ANOVA for slope

```
#table(slope)
aggregate(condition, by=list(slope),FUN=mean)
```

```
##   Group.1         x
## 1       0 0.2589928
## 2       1 0.6496350
## 3       2 0.5714286
```

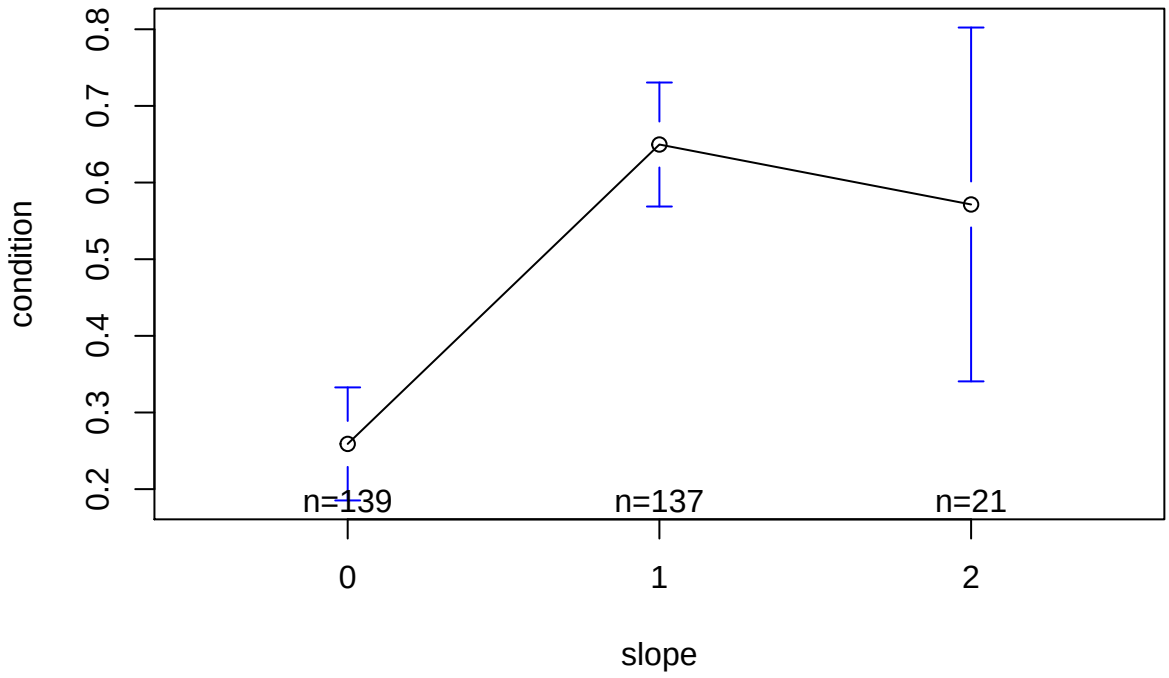
```r
aggregate(condition, by=list(slope),FUN=sd)
```

```
##   Group.1         x
## 1       0 0.4396660
## 2       1 0.4788350
## 3       2 0.5070926
```

```r
fitslope <- aov(condition ~ slope)
summary(fitslope)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## slope         1   8.19   8.187    36.8  4e-09 ***
## Residuals   295  65.62   0.222
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plotmeans(condition ~ slope)
```



one-way ANOVA for thal

```r
#table(thal)
aggregate(condition, by=list(thal),FUN=mean)
```

```
##   Group.1         x
## 1       0 0.2256098
## 2       1 0.6666667
## 3       2 0.7652174
```

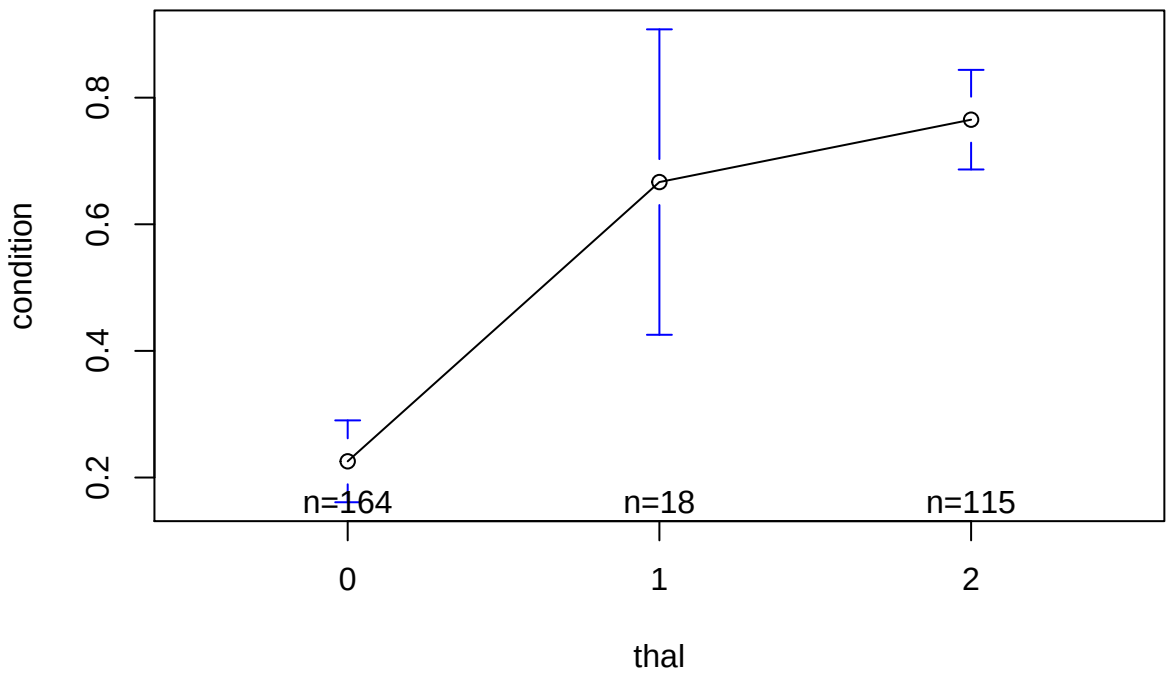
```r
aggregate(condition, by=list(thal),FUN=sd)
```

```
##   Group.1         x
## 1       0 0.4192634
## 2       1 0.4850713
## 3       2 0.4257179
```

```r
fitthal <- aov(condition ~ thal)
summary(fitthal)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## thal          1  20.00  19.996   109.6 <2e-16 ***
## Residuals   295  53.81   0.182
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plotmeans(condition ~ thal)
```



one-way ANOVA for ca

```
#table(ca)
aggregate(condition, by=list(ca),FUN=mean)
```

```
##   Group.1         x
## 1       0 0.2586207
## 2       1 0.6769231
## 3       2 0.8157895
## 4       3 0.8500000
```

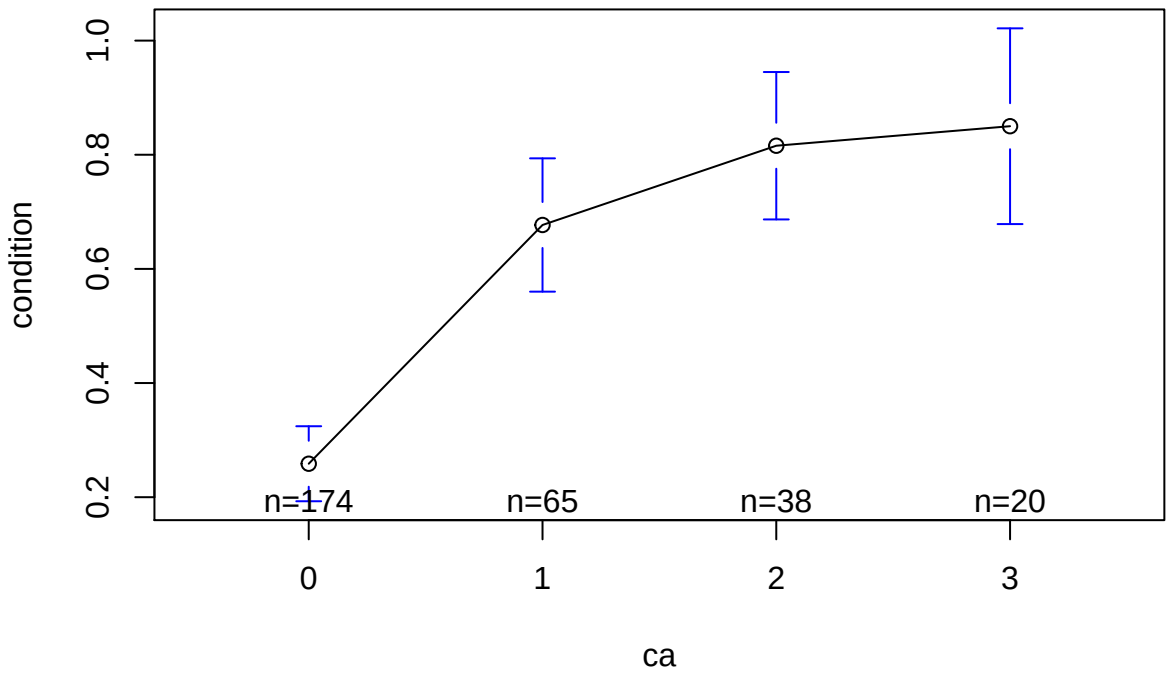
```
aggregate(condition, by=list(ca),FUN=sd)
```

```
##   Group.1         x
## 1       0 0.4391404
## 2       1 0.4712912
## 3       2 0.3928595
## 4       3 0.3663475
```

```
fitca <- aov(condition ~ ca)
summary(fitca)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## ca            1  15.83  15.834   80.58 <2e-16 ***
## Residuals   295  57.97   0.197
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plotmeans(condition ~ ca)
```



one-way ANOVA for sex

```
#table(sex)
aggregate(sex, by=list(condition),FUN=mean)
```

```
##   Group.1         x
## 1       0 0.5562500
## 2       1 0.8175182
```

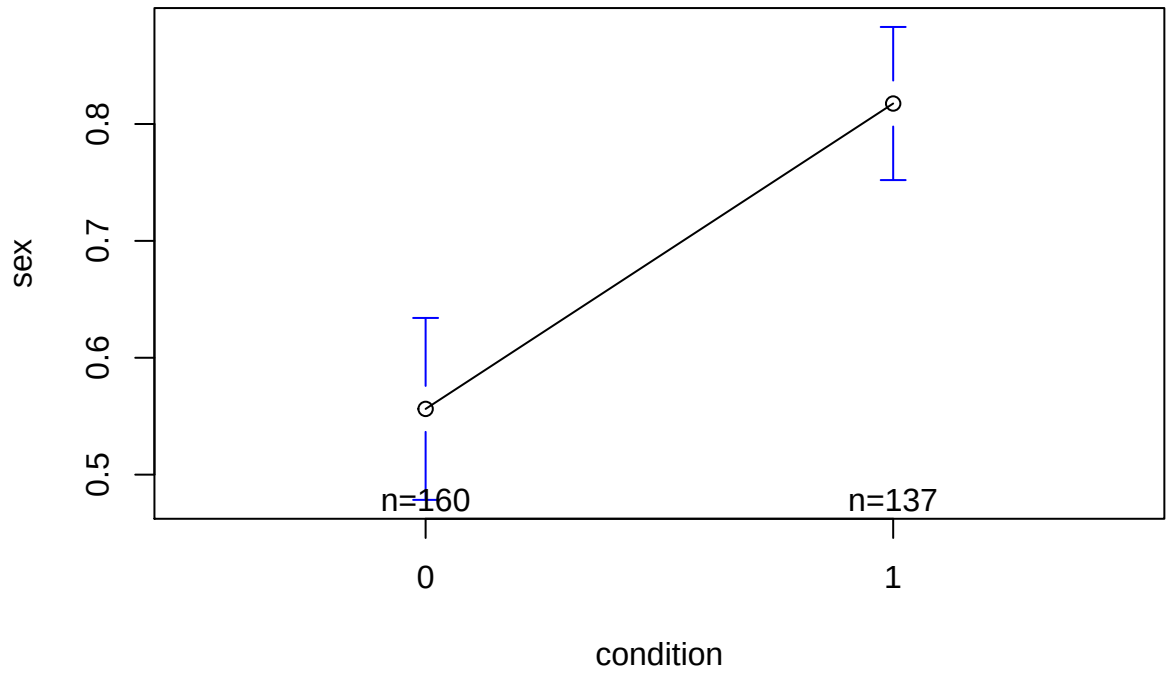
```
aggregate(sex, by=list(condition),FUN=sd)
```

```
##   Group.1         x
## 1       0 0.4983858
## 2       1 0.3876585
```

```
fitsex <- aov(sex ~ condition)
summary(fitsex)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## condition     1   5.04   5.038    24.8 1.09e-06 ***
## Residuals   295  59.93   0.203
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plotmeans(sex ~ condition)
```



one-way ANOVA for trestbps

```r
#table(trestbps)
aggregate(trestbps, by=list(condition),FUN=mean)
```

```
##   Group.1       x
## 1       0 129.175
## 2       1 134.635
```

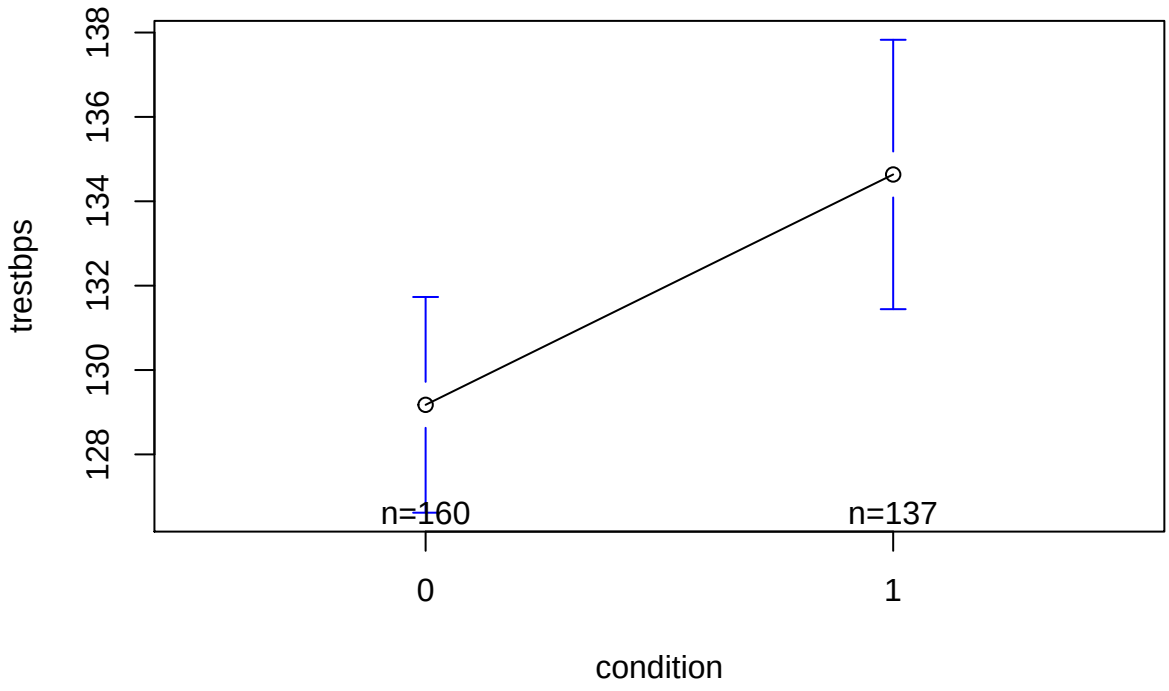
```r
aggregate(trestbps, by=list(condition),FUN=sd)
```

```
##   Group.1        x
## 1       0 16.37399
## 2       1 18.89673
```

```r
fittrestbps <- aov(trestbps ~ condition)
summary(fittrestbps)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## condition     1   2200  2200.3   7.118 0.00805 **
## Residuals   295  91193   309.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plotmeans(trestbps ~ condition)
```



one-way ANOVA for oldpeak

```
#table(oldpeak)
aggregate(oldpeak, by=list(condition),FUN=mean)
```

```
##   Group.1        x
## 1       0 0.598750
## 2       1 1.589051
```

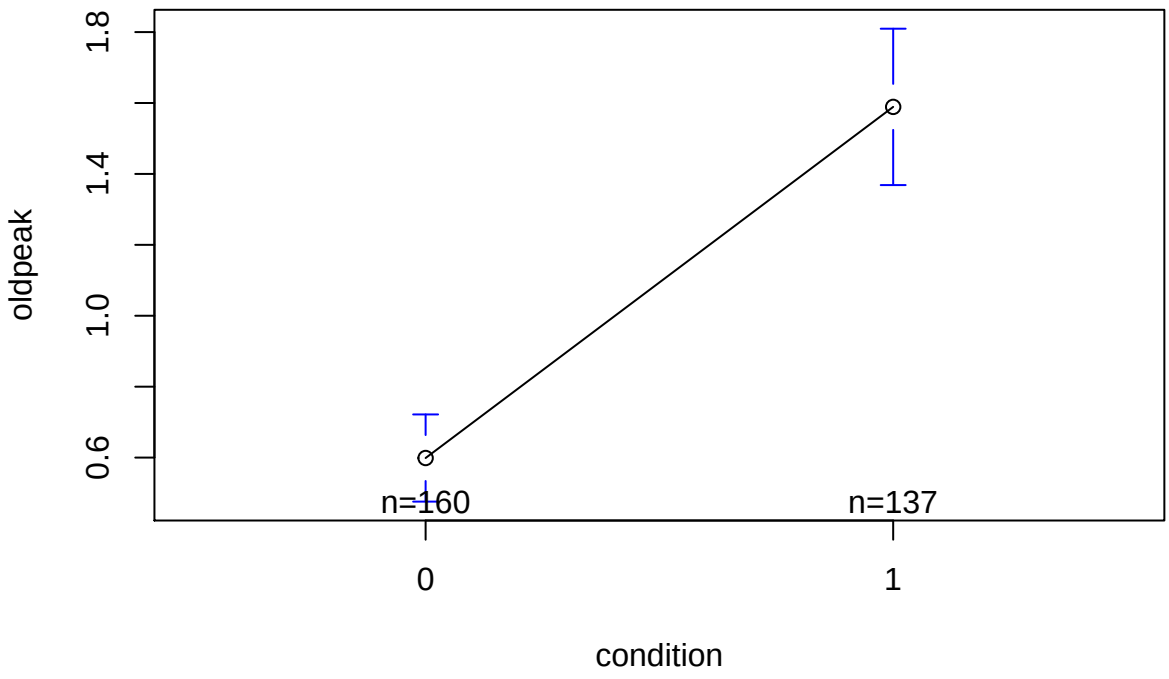
```
aggregate(oldpeak, by=list(condition),FUN=sd)
```

```
##   Group.1         x
## 1       0 0.7871601
## 2       1 1.3050061
```

```r
fitoldpeak <- aov(oldpeak ~ condition)
summary(fitoldpeak)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## condition     1   72.4   72.38   64.68 2.16e-14 ***
## Residuals   295  330.1    1.12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plotmeans(oldpeak ~ condition)
```



One-way ANCOVA for oldpeak and slope

```r
slope <- as.factor(slope)
fito_s <- aov(condition ~ oldpeak + slope)
summary(fito_s)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## oldpeak       1  13.27  13.272   68.64 4.29e-15 ***
## slope         2   3.88   1.939   10.03 6.14e-05 ***
## Residuals   293  56.66   0.193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```