

# Confidence intervals in R. ANOVA

Alla Tambovtseva, Olga Lyashevskaya, Ilya  
Schurov, gRik Moroz

09-02-2019

# Confidence intervals in R

Install library DescTools and load it:

```
# install.packages("DescTools")
```

```
library(DescTools)
```

Functions that we will use today:

- BinomCI -- binomial CI
- MeanCI -- CI for mean values
- MeanDiffCI -- CI for the difference in means

# Confidence intervals for proportions

First, let us consider an abstract example so as to look at different effects connected with confidence intervals (sample size effect and confidence level effect). Suppose we tossed a coin 20 times and got 4 heads.

```
nheads <- 4 # number of heads  
n1 <- 20 # total number of tosses
```

# Confidence intervals for proportions

First, let us consider an abstract example so as to look at different effects connected with confidence intervals (sample size effect and confidence level effect). Suppose we tossed a coin 20 times and got 4 heads.

```
nheads <- 4 # number of heads  
n1 <- 20 # total number of tosses
```

Now let's calculate a 95% confidence interval for the proportion of heads in such an experiment.

```
BinomCI(nheads, n1) # 95% by default
```

```
##      est      lwr.ci      upr.ci  
## [1,] 0.2 0.08065766 0.4160174
```

Calculate the length of a confidence interval:

```
ci.95 <- BinomCI(nheads, n1)  
ci.95[3] - ci.95[2]
```

```
## [1] 0.3353598
```

Calculate the length of a confidence interval:

```
ci.95 <- BinomCI(nheads, n1)
ci.95[3] - ci.95[2]
```

```
## [1] 0.3353598
```

Increase the number of tosses (number of heads remains the same):

```
n2 <- 40 # now 40 tosses
ci.95.2 <- BinomCI(nheads, n2)
ci.95.2[3] - ci.95.2[2]
```

```
## [1] 0.1909382
```

Calculate the length of a confidence interval:

```
ci.95 <- BinomCI(nheads, n1)
ci.95[3] - ci.95[2]
```

```
## [1] 0.3353598
```

Increase the number of tosses (number of heads remains the same):

```
n2 <- 40 # now 40 tosses
ci.95.2 <- BinomCI(nheads, n2)
ci.95.2[3] - ci.95.2[2]
```

```
## [1] 0.1909382
```

It shrinked, right?

Calculate the length of a confidence interval:

```
ci.95 <- BinomCI(nheads, n1)
ci.95[3] - ci.95[2]
```

```
## [1] 0.3353598
```

Increase the number of tosses (number of heads remains the same):

```
n2 <- 40 # now 40 tosses
ci.95.2 <- BinomCI(nheads, n2)
ci.95.2[3] - ci.95.2[2]
```

```
## [1] 0.1909382
```

It shrinked, right?

Keep the number of tosses equal to 20, but increase the confidence level:

```
ci.99 <- BinomCI(nheads, n1, conf.level = 0.99)
ci.99[3] - ci.99[2]
```

```
## [1] 0.4263411
```



Calculate the length of a confidence interval:

```
ci.95 <- BinomCI(nheads, n1)
ci.95[3] - ci.95[2]
```

```
## [1] 0.3353598
```

Increase the number of tosses (number of heads remains the same):

```
n2 <- 40 # now 40 tosses
ci.95.2 <- BinomCI(nheads, n2)
ci.95.2[3] - ci.95.2[2]
```

```
## [1] 0.1909382
```

It shrinked, right?

Keep the number of tosses equal to 20, but increase the confidence level:

```
ci.99 <- BinomCI(nheads, n1, conf.level = 0.99)
ci.99[3] - ci.99[2]
```

```
## [1] 0.4263411
```

It extended.

Let's set a true probability of getting a head in one toss of a coin.

```
p0 <- 0.5 # true probability of getting a head in
```

Then take 1000 samples of size 100, calculate confidence intervals for proportion of ones in each sample and count how many intervals contain a population proportion (the true probability of getting a head in one toss of a coin).

Recall the code from our previous seminars and suppose we asked 1000 people to toss a coin 100 times and report the proportion of heads they obtained.

```
n <- 100
samples <- 1000
dat <- matrix(sample(c(0, 1),
                    n * samples,
                    replace=TRUE), ncol=n, byrow
```

Let's set a true probability of getting a head in one toss of a coin.

```
p0 <- 0.5 # true probability of getting a head in
```

Then take 1000 samples of size 100, calculate confidence intervals for proportion of ones in each sample and count how many intervals contain a population proportion (the true probability of getting a head in one toss of a coin).

Recall the code from our previous seminars and suppose we asked 1000 people to toss a coin 100 times and report the proportion of heads they obtained.

```
n <- 100
samples <- 1000
dat <- matrix(sample(c(0, 1),
                     n * samples,
                     replace=TRUE), ncol=n, byrow
```

```
cis <- BinomCI(rowSums(dat), n)
colnames(cis) <- c("est", "lower", "upper")

head(cis[, "lower"])
```

```
##           x.1           x.2           x.3           x.4           x.5
## 0.3751082 0.3373330 0.4231658 0.4622810 0.4038315 0.38
```

```
head(cis[, "upper"])
```

```
##           x.1           x.2           x.3           x.4           x.5
## 0.5671114 0.5278461 0.6153545 0.6532797 0.5961685 0.57
```

```
sum(cis[,2] <= p0 & cis[,3] >= p0)
```

# Confidence intervals: real data

Now let's proceed to real data and work with *Verses* data set.

```
verses <- read.csv("https://raw.githubusercontent.com/johnfox77/verses_data/master/Verses.csv")
str(verses) # recall which variables are there
```

```
## 'data.frame':    364 obs. of  6 variables:
##  $ Decade          : Factor w/ 2 levels "1820s","1920s": 1 1 1 1 1 1 1 1 1 1 ...
##  $ RhymedNwords    : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ RhymedNsyl      : int  1 1 1 1 1 1 1 2 2 2 ...
##  $ UPoS            : Factor w/ 11 levels "ADJ","ADP","ADV","CONJ","DET","INTJ","NOUN","PRT","PRON","V","VBL": 1 1 1 1 1 1 1 1 1 1 ...
##  $ LineText        : Factor w/ 364 levels "-- Воронский, В.И.", "-- Воронский, В.И.", "-- Воронский, В.И.", "-- Воронский, В.И.", "-- Воронский, В.И.", "-- Воронский, В.И.", "-- Воронский, В.И.", "-- Воронский, В.И.", "-- Воронский, В.И.", "-- Воронский, В.И.", -- ...
##  $ Author           : Factor w/ 364 levels "А. А. Ахматова", "А. А. Ахматова", "А. А. Ахматова", "А. А. Ахматова", "А. А. Ахматова", "А. А. Ахматова", "А. А. Ахматова", "А. А. Ахматова", "А. А. Ахматова", "А. А. Ахматова", -- ...
```

# Confidence intervals: real data

Now let's proceed to real data and work with *Verses* data set.

```
verses <- read.csv("https://raw.githubusercontent.com/
str(verses) # recall which variables are there
```

```
## 'data.frame':   364 obs. of  6 variables:
## $ Decade       : Factor w/ 2 levels "1820s","1920s": 1
## $ RhymedNwords: int  1 1 1 1 1 1 1 1 1 1 ...
## $ RhymedNsyl   : int  1 1 1 1 1 1 1 2 2 2 ...
## $ UPoS         : Factor w/ 11 levels "ADJ","ADP","ADV"
## $ LineText     : Factor w/ 364 levels "-- Воронский, В
## $ Author       : Factor w/ 364 levels "А. А. Ахматова"
```

Calculate a confidence interval for the proportion of nouns at the end of lines:

```
nnouns <- nrow(verses[verses$UPoS == "NOUN", ])
total <- nrow(verses)

BinomCI(nnouns, total)
```

```
##           est      lwr.ci      upr.ci
## [1,] 0.6098901 0.5588825 0.6586025
```

# Confidence intervals for means

Now let's turn to the data set on Icelandic language from our previous class.

```
phono <- read.csv("http://math-info.hse.ru/f/2018
```

## Confidence intervals for means

Now let's turn to the data set on Icelandic language from our previous class.

```
phono <- read.csv("http://math-info.hse.ru/f/2018
```

Choose aspirated and non-aspirated cases again:

```
asp <- phono[phono$aspiration == "yes", ]  
nasp <- phono[phono$aspiration == "no", ]
```

# Confidence intervals for means

Now let's turn to the data set on Icelandic language from our previous class.

```
phono <- read.csv("http://math-info.hse.ru/f/2018
```

Choose aspirated and non-aspirated cases again:

```
asp <- phono[phono$aspiration == "yes", ]  
nasp <- phono[phono$aspiration == "no", ]
```

Calculate confidence intervals for mean values of vowel duration in each group:

```
MeanCI(asp$vowel.dur)
```

```
##      mean   lwr.ci   upr.ci  
## 78.75772 76.68274 80.83270
```

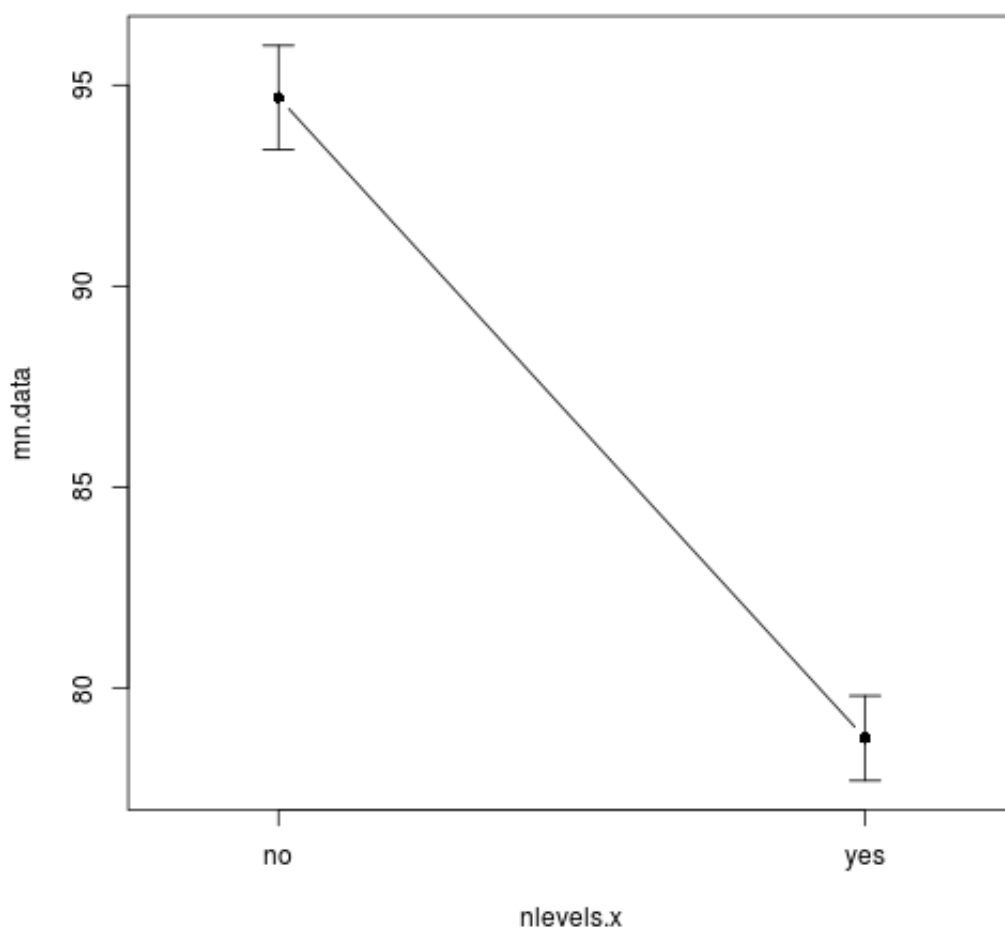
```
MeanCI(nasp$vowel.dur)
```

```
##      mean   lwr.ci   upr.ci  
## 94.69124 92.15292 97.22957
```



Plot them using `sciplot`:

```
# install.packages("sciplot")  
library(sciplot)  
lineplot.CI(data = phono,  
             response = vowel.dur,  
             x.factor = aspiration)
```



# Confidence intervals and significance of differences

- If two CI's for a population parameter (proportion, mean, median, etc) do not overlap, it means that true values of population parameters are significantly different.
- If two CI's for a population parameter overlap, true values of population parameters can (to be equal to each other), but **not** necessarily do so. For example, if two confidence intervals for means overlap, we cannot make a definite conclusion, more accurate testing is required (t-test). So, in general, comparison of confidence intervals (with the same confidence level, of course) is **not** equivalent to hypotheses testing.

# Confidence intervals and significance of differences

- If two CI's for a population parameter (proportion, mean, median, etc) do not overlap, it means that true values of population parameters are significantly different.
- If two CI's for a population parameter overlap, true values of population parameters can (to be equal to each other), but **not** necessarily do so. For example, if two confidence intervals for means overlap, we cannot make a definite conclusion, more accurate testing is required (t-test). So, in general, comparison of confidence intervals (with the same confidence level, of course) is **not** equivalent to hypotheses testing.

Consider a case when two CI's for means overlap, but population means are significantly different. Let's select only cases with aspirated consonants and compare the average vowel duration for round and unrounded vowels.

```
w1 <- phono[phono$aspiration == 'yes' & phono$rrou  
w2 <- phono[phono$aspiration == 'yes' & phono$rrou
```

## Do CI's overlap?

```
MeanCI(w1$vowel.dur)
```

```
##      mean   lwr.ci   upr.ci  
## 81.74052 77.89567 85.58537
```

```
MeanCI(w2$vowel.dur)
```

```
##      mean   lwr.ci   upr.ci  
## 76.90839 74.54499 79.27179
```

Can we conclude that mean vowel duration is different for round and unrounded vowels?

# T-test

Perform a statistical test, a two sample Student's t-test.

```
# reject or not reject H0  
t.test(w1$vowel.dur, w2$vowel.dur)
```

```
##  
##      Welch Two Sample t-test  
##  
## data:  w1$vowel.dur and w2$vowel.dur  
## t = 2.1134, df = 269.27, p-value = 0.03549  
## alternative hypothesis: true difference in means is not  
## 95 percent confidence interval:  
##  0.3304964 9.3337590  
## sample estimates:  
## mean of x mean of y  
##  81.74052  76.90839
```

# T-test

Perform a statistical test, a two sample Student's t-test.

```
# reject or not reject H0  
t.test(w1$vowel.dur, w2$vowel.dur)
```

```
##  
##      Welch Two Sample t-test  
##  
## data:  w1$vowel.dur and w2$vowel.dur  
## t = 2.1134, df = 269.27, p-value = 0.03549  
## alternative hypothesis: true difference in means is not  
## 95 percent confidence interval:  
##  0.3304964 9.3337590  
## sample estimates:  
## mean of x mean of y  
##  81.74052  76.90839
```

Null hypotheses should be rejected, so population means are different.

# T-test

Perform a statistical test, a two sample Student's t-test.

```
# reject or not reject H0  
t.test(w1$vowel.dur, w2$vowel.dur)
```

```
##  
##      Welch Two Sample t-test  
##  
## data:  w1$vowel.dur and w2$vowel.dur  
## t = 2.1134, df = 269.27, p-value = 0.03549  
## alternative hypothesis: true difference in means is not  
## 95 percent confidence interval:  
##  0.3304964 9.3337590  
## sample estimates:  
## mean of x mean of y  
##  81.74052  76.90839
```

Null hypotheses should be rejected, so population means are different.

Actually, testing hypothesis about the equality of population means is equivalent to finding whether *a CI for the difference of means* includes zero.

```
# CI for difference between means  
MeanDiffCI(w1$vowel.dur, w2$vowel.dur)
```

```
## meandiff      lwr.ci      upr.ci  
## 4.8321277 0.3304964 9.3337590
```

# T-test

Perform a statistical test, a two sample Student's t-test.

```
# reject or not reject H0
t.test(w1$vowel.dur, w2$vowel.dur)
```

```
##
##      Welch Two Sample t-test
##
## data:  w1$vowel.dur and w2$vowel.dur
## t = 2.1134, df = 269.27, p-value = 0.03549
## alternative hypothesis: true difference in means is not
## 95 percent confidence interval:
##  0.3304964 9.3337590
## sample estimates:
## mean of x mean of y
##  81.74052  76.90839
```

Null hypotheses should be rejected, so population means are different.

Actually, testing hypothesis about the equality of population means is equivalent to finding whether *a CI for the difference of means* includes zero.

```
# CI for difference between means
MeanDiffCI(w1$vowel.dur, w2$vowel.dur)
```

```
## meandiff      lwr.ci      upr.ci
## 4.8321277 0.3304964 9.3337590
```

Thus, intersection of CI's for means (or for any population parameters)  $\neq$  CI for the difference includes zero  $\neq H_0$  about equality should not be rejected.

11 / 20



# ANOVA

Load data on Icelandic:

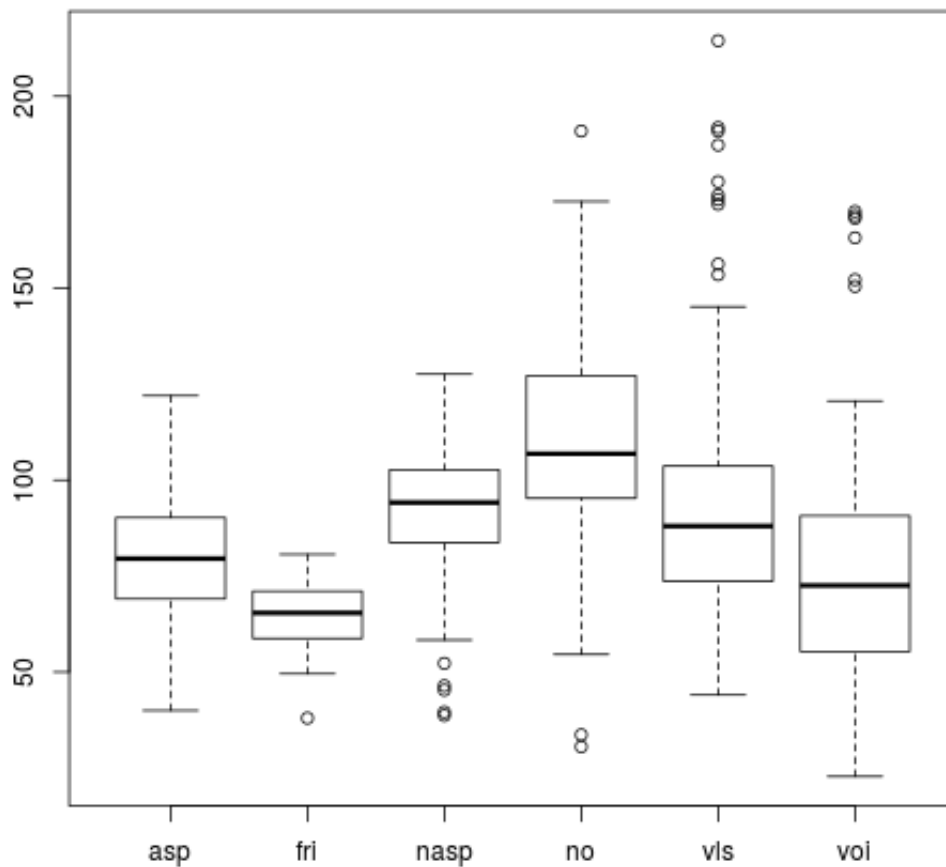
```
phono <- read.csv("http://math-info.hse.ru/f/2018
```

```
str(phono)
```

```
## 'data.frame':      806 obs. of  31 variables:
## $ speaker      : Factor w/  5 levels "brs02","bte03"
## $ index        : int   137 138 139 15 16 17 4 5 6 11
## $ word         : Factor w/ 58 levels "detta","dögg"
## $ time         : num   438 441 443 133 138 ...
## $ word.dur     : num   489 444 450 530 515 ...
## $ voicing.dur  : num   153.2 138.7 169.8 93.6 159.3
## $ vowel.dur    : num   100.9 72.6 107.5 93 107.4 ...
## $ cluster.dur  : num   231 217 195 245 230 ...
## $ spreading.dur : num   114.3 98.2 50.1 119.2 45.1 ..
## $ sonorant.dur : num   167 164 112 120 97 ...
## $ closure.dur  : num   64.7 52.7 82.3 125.3 133.2 ..
## $ vor          : num   332 290 302 338 338 ...
## $ voffr        : num   179 151 132 244 178 ...
## $ mor          : num   231 217 195 245 230 ...
## $ cond_no      : int    1 1 1 1 1 1 1 1 1 1 ...
## $ ipa          : Factor w/ 58 levels "celta","celta"
## $ cons1        : Factor w/  6 levels "asp","fri","na"
## $ vowel        : Factor w/ 10 levels "a","ai","e",""
## $ height       : Factor w/  4 levels "diph","high",.
## $ anteroposterior: Factor w/  2 levels "back","front":
## $ roundness    : Factor w/  2 levels "round","unround"
## $ consonant     : Factor w/ 18 levels "ʰk","ʰp","ʰt"
## $ manner       : Factor w/  5 levels "geminate","lat"
## $ place        : Factor w/  3 levels "coronal","labi"
## $ aspiration    : Factor w/  2 levels "no","yes"; 2 2
```

Create a boxplot for vowel duration for each group of consonants:

```
boxplot(phono$vowel.dur ~ phono$cons1)
```



Perform ANOVA to find out if the variation between groups is significantly larger than the variation within groups:

```
res <- aov(phono$vowel.dur ~ phono$cons1)
res
```

```
## Call:
##      aov(formula = phono$vowel.dur ~ phono$cons1)
##
## Terms:
##                phono$cons1 Residuals
## Sum of Squares      96776.3  404073.9
## Deg. of Freedom           5      800
##
## Residual standard error: 22.47426
## Estimated effects may be unbalanced
```

This type of ANOVA is called a one-way ANOVA for independent groups.

Perform ANOVA to find out if the variation between groups is significantly larger than the variation within groups:

```
res <- aov(phono$vowel.dur ~ phono$cons1)
res
```

```
## Call:
##      aov(formula = phono$vowel.dur ~ phono$cons1)
##
## Terms:
##              phono$cons1 Residuals
## Sum of Squares      96776.3  404073.9
## Deg. of Freedom           5      800
##
## Residual standard error: 22.47426
## Estimated effects may be unbalanced
```

This type of ANOVA is called a one-way ANOVA for independent groups. More informative summary:

```
# H0: there are no difference in population means
summary(res)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## phono$cons1    5  96776   19355   38.32 <2e-16 ***
## Residuals    800 404074     505
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0
```

## ANOVA with multiple groups:

```
res <- aov(phono$vowel.dur ~ phono$cons1 + phono$roundness)
res
```

```
## Call:
## aov(formula = phono$vowel.dur ~ phono$cons1 + phono$roundness, data = phono)
##
## Terms:
##              phono$cons1 phono$roundness Residuals
## Sum of Squares      96776.3          2037.5  402036.4
## Deg. of Freedom           5              1       799
##
## Residual standard error: 22.43155
## Estimated effects may be unbalanced
```

## ANOVA with multiple groups:

```
res <- aov(phono$vowel.dur ~ phono$cons1 + phono$roundness)
```

```
## Call:
## aov(formula = phono$vowel.dur ~ phono$cons1 + phono$roundness, data = phono)
##
## Terms:
##              phono$cons1 phono$roundness Residuals
## Sum of Squares      96776.3          2037.5  402036.4
## Deg. of Freedom           5              1       799
##
## Residual standard error: 22.43155
## Estimated effects may be unbalanced
```

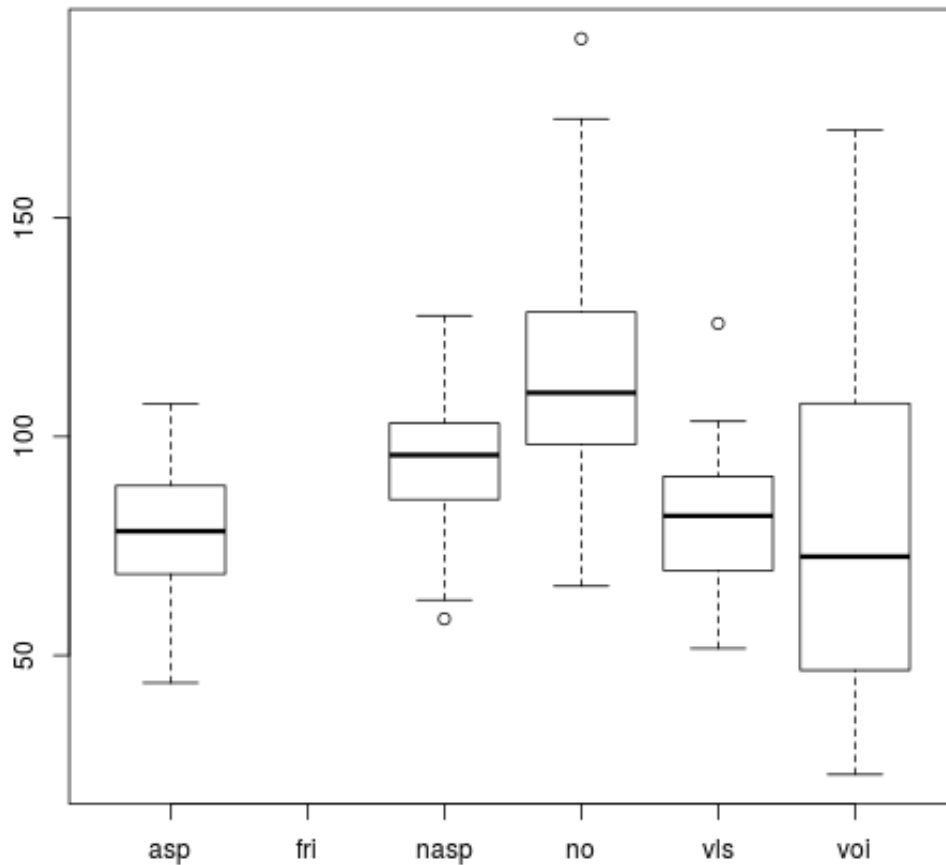
## More informative summary:

```
# H0: there are no difference in population means
summary(res)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## phono$cons1      5  96776    19355  38.466 <2e-16 ***
## phono$roundness  1   2037     2037   4.049  0.0445 *
## Residuals      799 402036     503
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

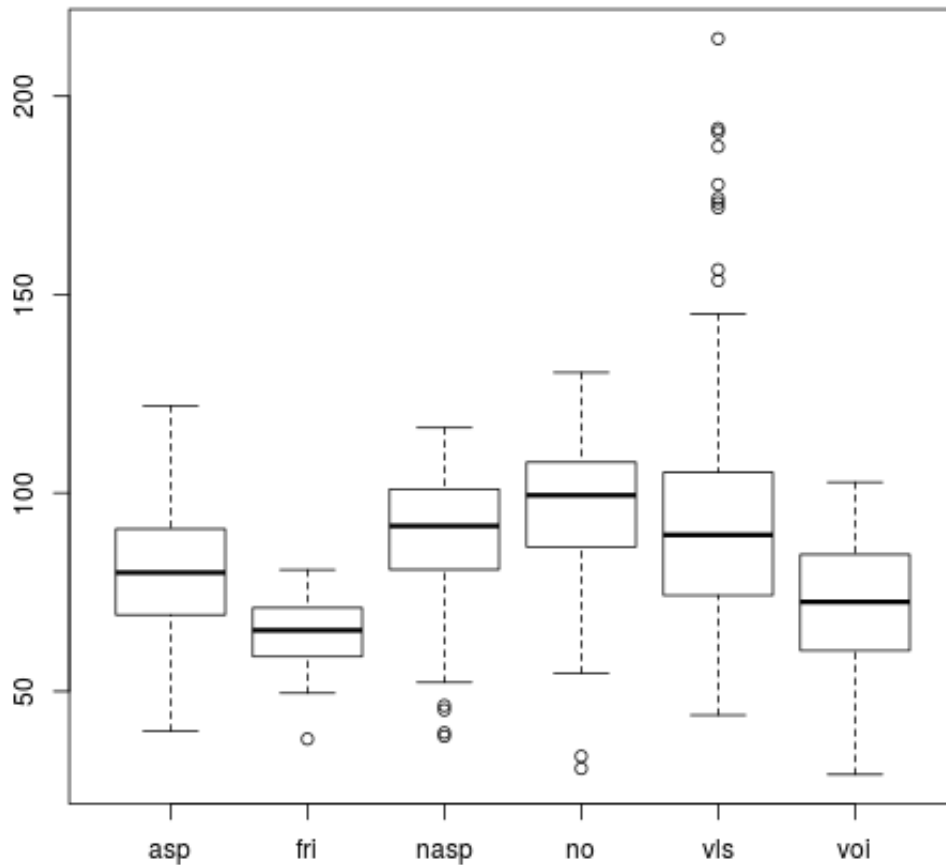
Create a boxplot for vowel duration for each group of consonants and for both groups of roundness:

```
boxplot(phono[phono$roundness == "round",]$vowel.
```



Create a boxplot for vowel duration for each group of consonants and for both groups of roundness:

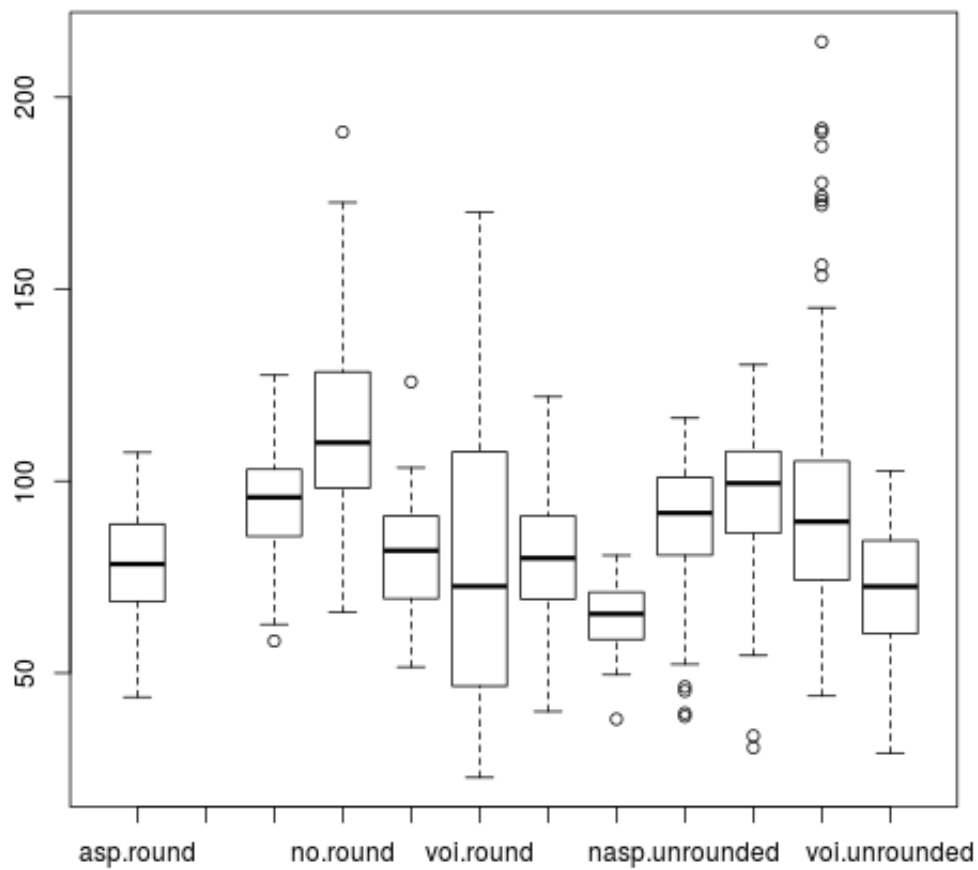
```
boxplot(phono[phono$roundness == "unrounded",]$vo
```





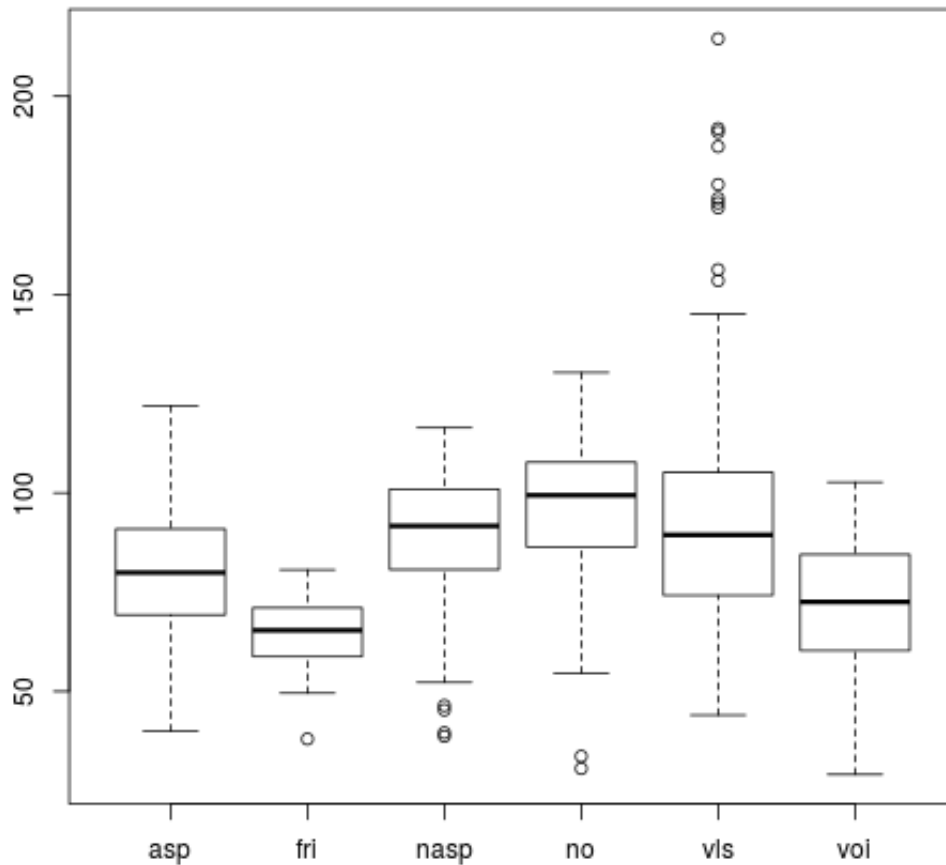
Plot all groups together:

```
boxplot(phono$vowel.dur ~ phono$cons1 + phono$rou
```



Create a boxplot for vowel duration for each group of consonants and for both groups of roundness:

```
boxplot(phono[phono$roundness == "unrounded",]$vo
```



## Create the same boxplot with ggplot2

```
ggplot(data = phono, # add the data
       aes(x = cons1, y = vowel.dur, # set x, y c
           color = cons1)) +      # color by cons1
  geom_boxplot() +
  facet_grid(~roundness) # create panes base on h
```

