ANOVA Implementation in R

18BCE1104 - Ankita Duraphe

18/02/2021

### ANOVA implementation on penguins dataset

1. What is ANOVA? - ANalysis Of VAriance- statistical test - 2 or more pop means are different - 2 or more groups are significantly diff. An ANOVA test is a way to find out if survey or experiment results are significant.
2. Why the name or how it works? - ANOVA (means) - compares “between” variance and the variance “within” - If the between variance is significantly larger than the within variance,group means are declared to be diff between variance/within variance - by comparing this ratio to a threshold from the Fisher probability distribution (5%)
3. Similar test?! - student t-test (2) - ANOVA (3 or more)
4. What is continuous variable?
5. What is Categorical variable?
6. Why ANOVA ? - investigate relations between categorical variable and continuous variable
7. ANOVA used for? - type of hypothesis testing for population variance
8. Types of ANOVA - 1-way, 2-way & sometimes 3-way ANOVA, mixed ANOVA, repeated measures ANOVA etc…

library(palmerpenguins)  
library(tidyverse)

View the dataset

View(penguins)

selecting features from dataset

data = penguins[, c("species", "flipper\_length\_mm")]

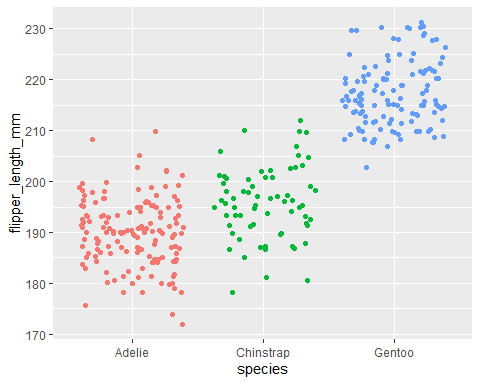
some descriptive statistics

summary(data)

## species flipper\_length\_mm  
## Adelie :152 Min. :172.0   
## Chinstrap: 68 1st Qu.:190.0   
## Gentoo :124 Median :197.0   
## Mean :200.9   
## 3rd Qu.:213.0   
## Max. :231.0   
## NA's :2

data visulaization

library(ggplot2)  
ggplot(data) +  
 aes(x = species, y = flipper\_length\_mm, color = species) +  
 geom\_jitter() +  
 theme(legend.position = "none")



Aim of ANOVA:

1. Measurements (mean, variance) are similar across groups of categorical variable (here species).
2. Compare impact of diff levels of categorical variable on a quantitaive variable (flippers length).
3. Explain quantitative variable based on qualitative variable.

“Are flippers length different for the 3 species of penguins?”

Null and Alternative hypothesis of an ANOVA:

1. H0: ??\_Adelie = ??\_Chinstrap = ??\_Gentoo (??? the 3 species are equal in terms of flipper length)
2. H1: at least one mean is different (??? at least one species is different from the other 2 species in terms of flipper length)

Assumptions of ANOVA:

1. Check that your observations are independent (Durbin-Watson test)
2. If they are independent, test the normality of residuals,
   1. If normality is assumed, test the homogeneity of the variances:
   2. If variances are equal, use ANOVA.
   3. If variances are not equal, use the Welch test.
   4. If normality is not assumed, use the Kruskal-Wallis test.

ANOVA evaluation - residuals

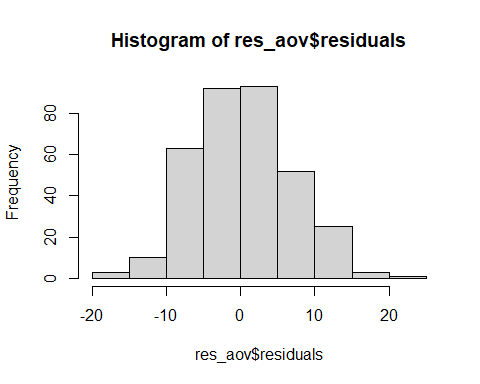
res\_aov = aov(flipper\_length\_mm ~ species, data = data)

Visual way of Normality check (histogram / QQ plot)

par(mfrow = c(1, 2))

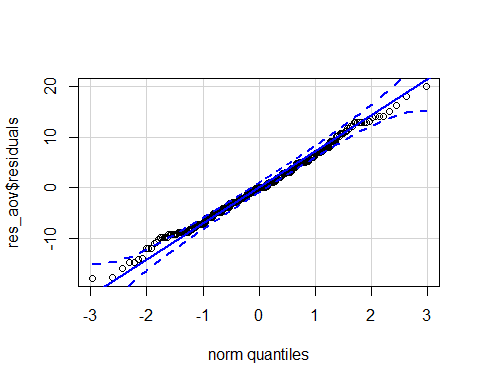
histogram

hist(res\_aov$residuals)



QQ-plot (within the confidence bands?)

library(car)  
qqPlot(res\_aov$residuals,id = FALSE)



Note: ANOVA is quite robust to small deviations from normality

Formal way of Normality check

shapiro.test(res\_aov$residuals)

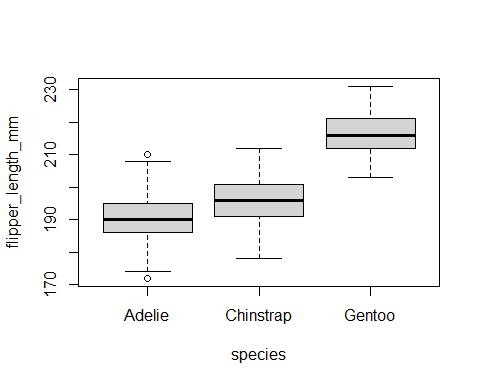
##   
## Shapiro-Wilk normality test  
##   
## data: res\_aov$residuals  
## W = 0.99452, p-value = 0.2609

P value is larger than 5% - normal distribution (residuals)  
p-value <?? (Null hypothesis gets rejected)

Equality of variances - homogeneity (Visually)

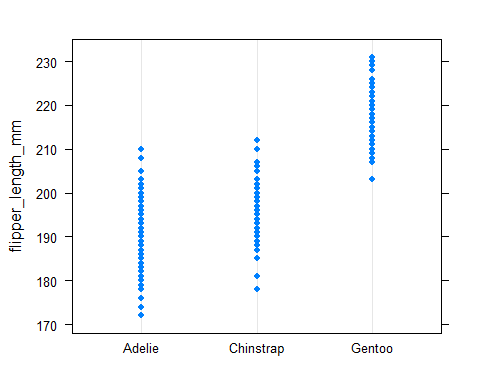
Boxplot

boxplot(flipper\_length\_mm ~ species, data = data)



Dotplot

library("lattice")  
dotplot(flipper\_length\_mm ~ species, data = data)



Formal way - Levene’s test

library(car)  
leveneTest(flipper\_length\_mm ~ species,  
 data = data)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 2 0.3306 0.7188  
## 339

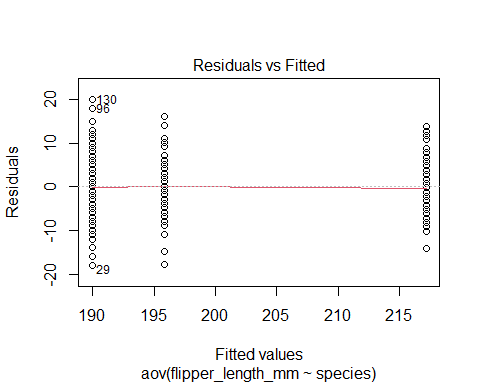
p-value being larger than the significance level of 0.05 - null hypothesis

Another method to test normality and homogeneity

par(mfrow = c(1, 2))

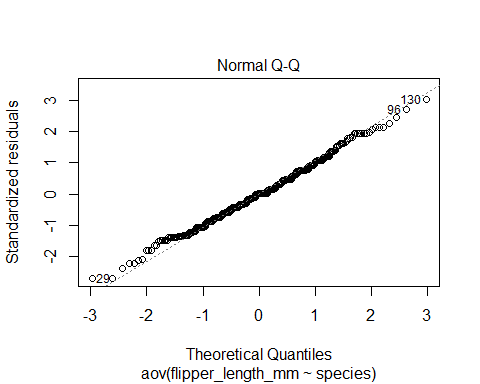
1. Homogeneity of variances (red line is flat)

plot(res\_aov, which = 1)



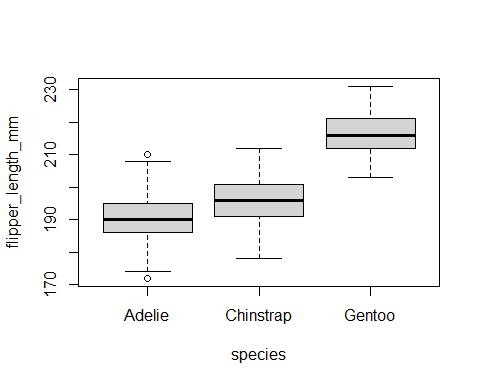
1. Normality (approx follow dot line)

plot(res\_aov, which = 2)



ANOVA - Preliminary analyses

boxplot(flipper\_length\_mm ~ species,  
 data = data  
)



descriptive statistics

aggregate(flipper\_length\_mm ~ species,  
 data = data,  
 function(x) round(c(mean = mean(x), sd = sd(x)), 2)  
)

## species flipper\_length\_mm.mean flipper\_length\_mm.sd  
## 1 Adelie 189.95 6.54  
## 2 Chinstrap 195.82 7.13  
## 3 Gentoo 217.19 6.48

not enough to conclude that flippers are significantly different in the 3 populations of penguins

ANOVA

1st method:  
advantage of the first method is that it is easy to switch from the ANOVA (used when variances are equal)

oneway.test(flipper\_length\_mm ~ species,  
 data = data,  
 var.equal = TRUE # assuming equal variances  
)

##   
## One-way analysis of means  
##   
## data: flipper\_length\_mm and species  
## F = 594.8, num df = 2, denom df = 339, p-value < 2.2e-16

to the Welch test (used when variances are unequal)

oneway.test(flipper\_length\_mm ~ species,  
 data = data,  
 var.equal = FALSE # assuming unequal variances  
)

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: flipper\_length\_mm and species  
## F = 614.01, num df = 2.00, denom df = 172.76, p-value < 2.2e-16

2nd method:

res\_aov = aov(flipper\_length\_mm ~ species,  
 data = data  
)  
summary(res\_aov)

## Df Sum Sq Mean Sq F value Pr(>F)   
## species 2 52473 26237 594.8 <2e-16 \*\*\*  
## Residuals 339 14953 44   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## 2 observations deleted due to missingness

Interpreataion

1. p-value is smaller than 0.05 - reject null hypothesis
2. at least one species is different than the others in terms of flippers length (p-value < 2.2e-16)

Report results of ANOVA

# remotes::install\_github("easystats/report")

Results of an ANOVA, however, do NOT tell us which group(s) is(are) different from the others  
- post-hoc tests - compare groups 2 by 2 :-  
Turkey HSD / Dunnetts test  
Chinstrap versus Adelie  
Gentoo vs. Adelie  
Gentoo vs. Chinstrap

library("report") # Load the package every time you start R  
report(res\_aov)

## The ANOVA (formula: flipper\_length\_mm ~ species) suggests that:  
##   
## - The main effect of species is statistically significant and large (F(2, 339) = 594.80, p < .001; Eta2 = 0.78, 90% CI [0.75, 0.80])  
##   
## Effect sizes were labelled following Field's (2013) recommendations.

library(multcomp)

Tukey HSD test: 2nd method ANOVA is used here

post\_test = glht(res\_aov,  
 linfct = mcp(species = "Tukey")  
)  
summary(post\_test)

##   
## Simultaneous Tests for General Linear Hypotheses  
##   
## Multiple Comparisons of Means: Tukey Contrasts  
##   
##   
## Fit: aov(formula = flipper\_length\_mm ~ species, data = data)  
##   
## Linear Hypotheses:  
## Estimate Std. Error t value Pr(>|t|)   
## Chinstrap - Adelie == 0 5.8699 0.9699 6.052 6.87e-09 \*\*\*  
## Gentoo - Adelie == 0 27.2333 0.8067 33.760 < 1e-09 \*\*\*  
## Gentoo - Chinstrap == 0 21.3635 1.0036 21.286 < 1e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## (Adjusted p values reported -- single-step method)

All three p-values are smaller than 0.05

library(multcomp)

Dunnett’s test: (Adelie is reference)

post\_test <- glht(res\_aov,  
 linfct = mcp(species = "Dunnett")  
)  
summary(post\_test)

##   
## Simultaneous Tests for General Linear Hypotheses  
##   
## Multiple Comparisons of Means: Dunnett Contrasts  
##   
##   
## Fit: aov(formula = flipper\_length\_mm ~ species, data = data)  
##   
## Linear Hypotheses:  
## Estimate Std. Error t value Pr(>|t|)   
## Chinstrap - Adelie == 0 5.8699 0.9699 6.052 7.59e-09 \*\*\*  
## Gentoo - Adelie == 0 27.2333 0.8067 33.760 < 1e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## (Adjusted p values reported -- single-step method)

Change reference category: instead of adelie now gentoo

data$species <- relevel(data$species, ref = "Gentoo")

Check that Gentoo is the reference category:

levels(data$species)

## [1] "Gentoo" "Adelie" "Chinstrap"

res\_aov2 <- aov(flipper\_length\_mm ~ species,  
 data = data  
)

Dunnett’s test:

post\_test <- glht(res\_aov2,  
 linfct = mcp(species = "Dunnett")  
)  
summary(post\_test)

##   
## Simultaneous Tests for General Linear Hypotheses  
##   
## Multiple Comparisons of Means: Dunnett Contrasts  
##   
##   
## Fit: aov(formula = flipper\_length\_mm ~ species, data = data)  
##   
## Linear Hypotheses:  
## Estimate Std. Error t value Pr(>|t|)   
## Adelie - Gentoo == 0 -27.2333 0.8067 -33.76 <1e-10 \*\*\*  
## Chinstrap - Gentoo == 0 -21.3635 1.0036 -21.29 <1e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## (Adjusted p values reported -- single-step method)

### One-way ANOVA for Iris Dataset

**Checking Assumptions**

The assumptions of ANOVA are:

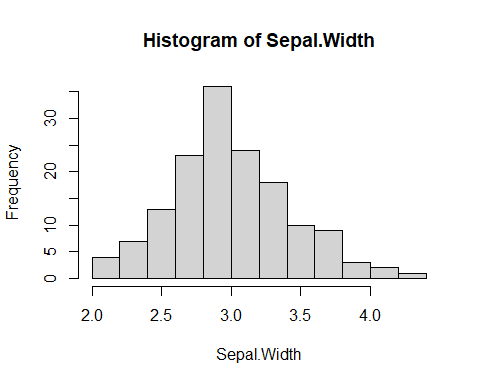
1. Samples are independent from one another (within and among groups).
2. Samples were randomly chosen.
3. Response variable is normally distributed.
4. Variances are equal among groups.

You have to determine if your sampling was independent and random. The other assumptions can be tested with the code below.

**Testing Normality**

First we will test normality using two methods. The first is by looking at the histogram using the hist() function. The second way is to use a statistical test for normality called the Shapiro-Wilk test.

attach(iris)   
hist(Sepal.Width)



The histogram is rather normal, but maybe has a bit of right skew. The Shapiro-Wilk test shows a p-value of 0.10, which is greater than 0.05, suggesting that the data is normally distributed. In summary, if the Shapiro-Wilk p-value is greater than 0.05, it suggests the data is normal; if it’s less than 0.05, then it’s nonnormal.

shapiro.test(Sepal.Width)

##   
## Shapiro-Wilk normality test  
##   
## data: Sepal.Width  
## W = 0.98492, p-value = 0.1012

**Testing Variances**

Next we will test if the variances are the same among the groups using the Levene’s test.

We will call the package car to run the function leveneTest.

library(car)   
lm.iris1<-lm(Sepal.Width~Species,data=iris)   
leveneTest(lm.iris1)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 2 0.5902 0.5555  
## 147

The variances are equal, as the p-value is 0.556 (greater than 0.05).

This dataset satisfies our assumptions, so we can run the ANOVA!

**Running the one-way ANOVA**

anova(lm.iris1)

## Analysis of Variance Table  
##   
## Response: Sepal.Width  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Species 2 11.345 5.6725 49.16 < 2.2e-16 \*\*\*  
## Residuals 147 16.962 0.1154   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

There is at least one group that is significantly different (p is super small: p< 2.2 x 10^-16).

The summary gives the average of the first species, setosa (estimate of the intercept) and how the other groups are different from setosa (see the estimates of Speciesversicolor and Speciesvirginica). I. ersicolor has, on average, sepals that are 0.658 mm narrower that I. setosa; I. virginica is 0.454 mm narrower than I. setosa.

summary(lm.iris1)

##   
## Call:  
## lm(formula = Sepal.Width ~ Species, data = iris)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.128 -0.228 0.026 0.226 0.972   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.42800 0.04804 71.359 < 2e-16 \*\*\*  
## Speciesversicolor -0.65800 0.06794 -9.685 < 2e-16 \*\*\*  
## Speciesvirginica -0.45400 0.06794 -6.683 4.54e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3397 on 147 degrees of freedom  
## Multiple R-squared: 0.4008, Adjusted R-squared: 0.3926   
## F-statistic: 49.16 on 2 and 147 DF, p-value: < 2.2e-16

**Post-hoc Multiple Comparisons**

In the first line of code below, we create another ANOVA object using aov(). This is very similar to using lm() and anova(), like we did above. We use aov() below because it is compatible with the function TukeyHSD() in the second line of code.

iris.aov<-aov(Sepal.Width~Species,data=iris)  
TukeyHSD(iris.aov)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = Sepal.Width ~ Species, data = iris)  
##   
## $Species  
## diff lwr upr p adj  
## versicolor-setosa -0.658 -0.81885528 -0.4971447 0.0000000  
## virginica-setosa -0.454 -0.61485528 -0.2931447 0.0000000  
## virginica-versicolor 0.204 0.04314472 0.3648553 0.0087802

Look at the resulting Tukey table. You should see 4 columns and 3 rows. Each row is a comparison between 2 of the species. Look at the 4th column, p adj. This is the p-value for the comparison; less than 0.05 the two groups are different, greater than 0.05 and they are not significanlty different from one another.

### Two-way ANOVA for Iris Dataset

Now we will do a two-way ANOVA using the iris dataset again. This time, we will predict if sepal width differs among species and community.

The next code makes the community variable in the iris dataset.

v<-c("high","low")  
iris$community<-v

**Checking Assumptions**

Can check normality using the same code above

**Check variances**

We will again use a Levene test.

library(car)   
lm.iris2<-lm(Sepal.Width~Species\*community,data=iris)   
leveneTest(lm.iris2)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 5 0.3307 0.8938  
## 144

The Levene p-value is greater than 0.05

**Running the two-way ANOVA**

anova(lm.iris2)

## Analysis of Variance Table  
##   
## Response: Sepal.Width  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Species 2 11.3449 5.6725 48.7581 <2e-16 \*\*\*  
## community 1 0.0067 0.0067 0.0573 0.8111   
## Species:community 2 0.2025 0.1013 0.8704 0.4210   
## Residuals 144 16.7528 0.1163   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Look at the anova table. The farthest right column (Pr(>F)) shows a p-value of 0.42, the interaction of the 2 variables is not significant. Now in the community row, we see a p-value of 0.81 - community does not have a significant effect on sepal width. Finally, we see that Species has a super small p-value (less than 2 X 10^-16). Species of iris has a significant effect on sepal width.

**Post-hoc comparisons**

Because the only signicant factor was Species, this post-hoc comparison will be the same as the one we did earlier.

iris2.aov<-aov(Sepal.Width~Species,data=iris)  
TukeyHSD(iris2.aov)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = Sepal.Width ~ Species, data = iris)  
##   
## $Species  
## diff lwr upr p adj  
## versicolor-setosa -0.658 -0.81885528 -0.4971447 0.0000000  
## virginica-setosa -0.454 -0.61485528 -0.2931447 0.0000000  
## virginica-versicolor 0.204 0.04314472 0.3648553 0.0087802

iris3.aov<-aov(Sepal.Width~Species\*community,data=iris)  
TukeyHSD(iris3.aov)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = Sepal.Width ~ Species \* community, data = iris)  
##   
## $Species  
## diff lwr upr p adj  
## versicolor-setosa -0.658 -0.81955127 -0.4964487 0.0000000  
## virginica-setosa -0.454 -0.61555127 -0.2924487 0.0000000  
## virginica-versicolor 0.204 0.04244873 0.3655513 0.0091385  
##   
## $community  
## diff lwr upr p adj  
## low-high -0.01333333 -0.1234264 0.09675978 0.8111496  
##   
## $`Species:community`  
## diff lwr upr p adj  
## versicolor:high-setosa:high -0.704 -0.98266057 -0.42533943 0.0000000  
## virginica:high-setosa:high -0.544 -0.82266057 -0.26533943 0.0000013  
## setosa:low-setosa:high -0.104 -0.38266057 0.17466057 0.8894028  
## versicolor:low-setosa:high -0.716 -0.99466057 -0.43733943 0.0000000  
## virginica:low-setosa:high -0.468 -0.74666057 -0.18933943 0.0000459  
## virginica:high-versicolor:high 0.160 -0.11866057 0.43866057 0.5613861  
## setosa:low-versicolor:high 0.600 0.32133943 0.87866057 0.0000001  
## versicolor:low-versicolor:high -0.012 -0.29066057 0.26666057 0.9999958  
## virginica:low-versicolor:high 0.236 -0.04266057 0.51466057 0.1475984  
## setosa:low-virginica:high 0.440 0.16133943 0.71866057 0.0001553  
## versicolor:low-virginica:high -0.172 -0.45066057 0.10666057 0.4800283  
## virginica:low-virginica:high 0.076 -0.20266057 0.35466057 0.9692150  
## versicolor:low-setosa:low -0.612 -0.89066057 -0.33333943 0.0000000  
## virginica:low-setosa:low -0.364 -0.64266057 -0.08533943 0.0031579  
## virginica:low-versicolor:low 0.248 -0.03066057 0.52666057 0.1113213

**t-test**

A t-test looks for a difference in means between two groups.

First we will create a new dataset that only has two species of iris using the subset() function.

iris.2<-subset(iris,Species!="virginica") #drops all virginica records

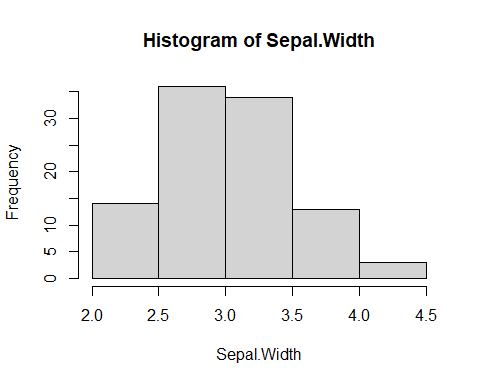
**Testing assumptions**

A t-test has almost the same assumptions as ANOVA.

**Checking normality**

attach(iris.2)

hist(Sepal.Width)



shapiro.test(Sepal.Width)

##   
## Shapiro-Wilk normality test  
##   
## data: Sepal.Width  
## W = 0.98977, p-value = 0.6463

The histogram has a bit of right skew. The Shapiro test p-value is above 0.05, so it’s normally distributed!

**Testing variances**

leveneTest(Sepal.Width,Species,data=iris.2)

## Levene's Test for Homogeneity of Variance (center = median: iris.2)  
## Df F value Pr(>F)  
## group 1 0.591 0.4439  
## 98

A Levene test gives a p-value of 0.44, above 0.05, so the variances are not different.

Note: if your variances ARE different, you can use a Welch’s t-test.

Second note: we used a little different format for the Levene test for this test. They are interchangeable: you could have made the lm (linear model) and ran the Levene’s test on it, like we did with the ANOVAs.

**Running the t-test**

t.test(Sepal.Width~Species,data=iris.2)

##   
## Welch Two Sample t-test  
##   
## data: Sepal.Width by Species  
## t = 9.455, df = 94.698, p-value = 2.484e-15  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.5198348 0.7961652  
## sample estimates:  
## mean in group setosa mean in group versicolor   
## 3.428 2.770

The resulting t-test table is more easily interpreted than most tables in R. It gives you a t-value, degrees of freedom, p-value, and means of both groups. Because the p-value is less than 0.05, sepal width is significantly different between the two species.

**Conclusion:**  
Basic ANOVA implementation has been performed on penguins dataset. Also, both One-way and Two-way ANOVA has been successfully performed on iris dataset.