

Proline distribution

mean(non\_proline\_angles)

# [1] 168.9382

sd(non\_proline\_angles)

# [1] 14.36636

#normality tests

ad.test(non\_proline\_angles)

# p-value < 0.00000000000000022

# Ho is that the data is normally distributed. Ho is rejected, the data does not have a normal distribution

ks.test(non\_proline\_angles,pnorm(10000,168.9382,14.36636))

#Two-sample Kolmogorov-Smirnov test

#data: non\_proline\_angles and pnorm(10000, 168.9382, 14.36636)

#D = 1, p-value = 0.27

#alternative hypothesis: two-sided

# Ho is that non\_proline\_anges is from a normal population with mean 168.9382 and sd 14.366. Ho is accepted as p\_val >0.05

#Shapiro Wilks test

non\_proline\_p\_values <- numeric()

for (i in 1:5000){

temp\_samp <- sample(5000,non\_proline\_angles)

test\_result <- shapiro.test(temp\_samp)

non\_proline\_p\_values <- c(non\_proline\_p\_values,test\_result$p.value)

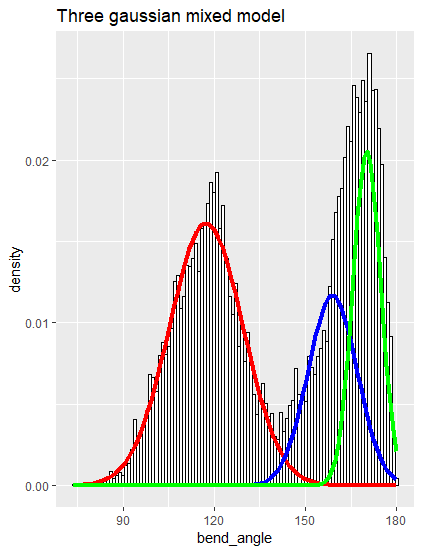
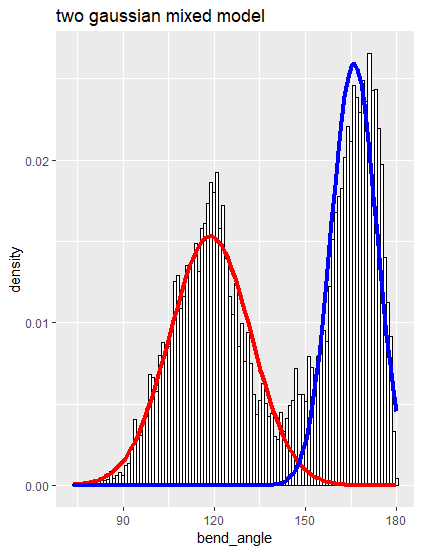
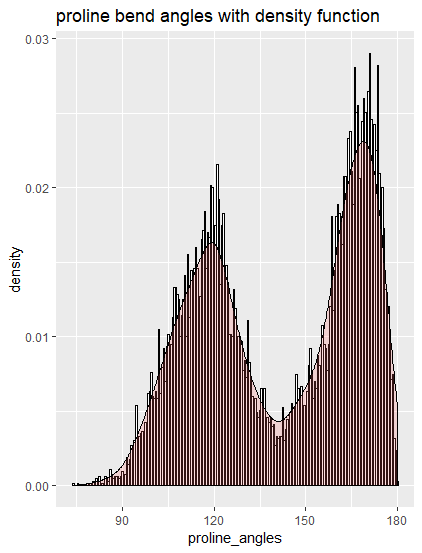
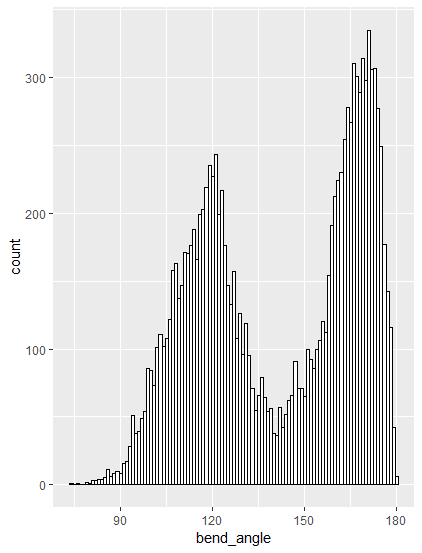
}

non\_proline\_p\_values[0.05<non\_proline\_p\_values]

# no pvalues >0.05 = 0

# Ho is that the data is normally distributed. Ho is rejected.

#Based on these results I would conclude that non\_proline data will have to be treated as non normal. This means we will require non parametric comparison.



Mixed Gaussian models

Two gaussian mixed model summary

summary of normalmixEM object:

comp 1 comp 2

lambda 0.513047 0.486953

mu 118.760184 165.911344

sigma 13.361777 7.496641

loglik at estimate: -55214.06

Three gaussian mixed model summary

summary of normalmixEM object:

comp 1 comp 2 comp 3

lambda 0.488383 0.235265 0.276352

mu 117.372331 158.892785 170.130773

sigma 12.093170 8.046908 4.571051

loglik at estimate: -54886.06

Analysis of results

mean(non\_proline\_angles)

# [1] 168.9382

sd(non\_proline\_angles)

# [1] 14.36636

#normality tests

ad.test(non\_proline\_angles)

# p-value < 0.00000000000000022

# not\_normal distribution

ks.test(non\_proline\_angles,pnorm(10000,168.9382,14.36636))

#Two-sample Kolmogorov-Smirnov test

#data: non\_proline\_angles and pnorm(10000, 168.9382, 14.36636)

#D = 1, p-value = 0.27

#alternative hypothesis: two-sided

# Ho is that both are from the same population. Ho is accepted as p\_val >0.05

#Shapiro Wilks test

non\_proline\_p\_values <- numeric()

for (i in 1:5000){

temp\_samp <- sample(5000,non\_proline\_angles)

test\_result <- shapiro.test(temp\_samp)

non\_proline\_p\_values <- c(non\_proline\_p\_values,test\_result$p.value)

}

non\_proline\_p\_values[0.05<non\_proline\_p\_values]

# no pvalues >0.05 = 0

# Ho is that the data is normally distributed. Ho is rejected.

# comparison tests between proline 3rd mixed gaussian model(of 3)

# and non\_proline data

# proline\_3rd mixed gaussian

# mean =170.130773 sd = 4.571051

proline\_3rd\_mixed\_sample <- rnorm(10000,mean = 170.130773,sd = 4.571051)

# non\_proline bend angles

# mean = 168.9382 sd = 14.3663

# this means we shall use unquel varience test

t.test(proline\_3rd\_mixed\_sample,non\_proline\_angles)

# Ho is that the difference in means is = to zero.

# p-value <0.00000000000000022

# we reject Ho