# Sai Lakshman Rao Duggirala – M15848995

#### Wine Project Part B – 1(e) , 2(d),3(d)

Question

1: Suppose the population mean of the variable “density” is μ, do the following-inferences:

1. Can we use a normal distribution to model “density”? If yes, what are the maximum likelihood estimates of the mean and standard deviation? Please provide their standard errors as well.

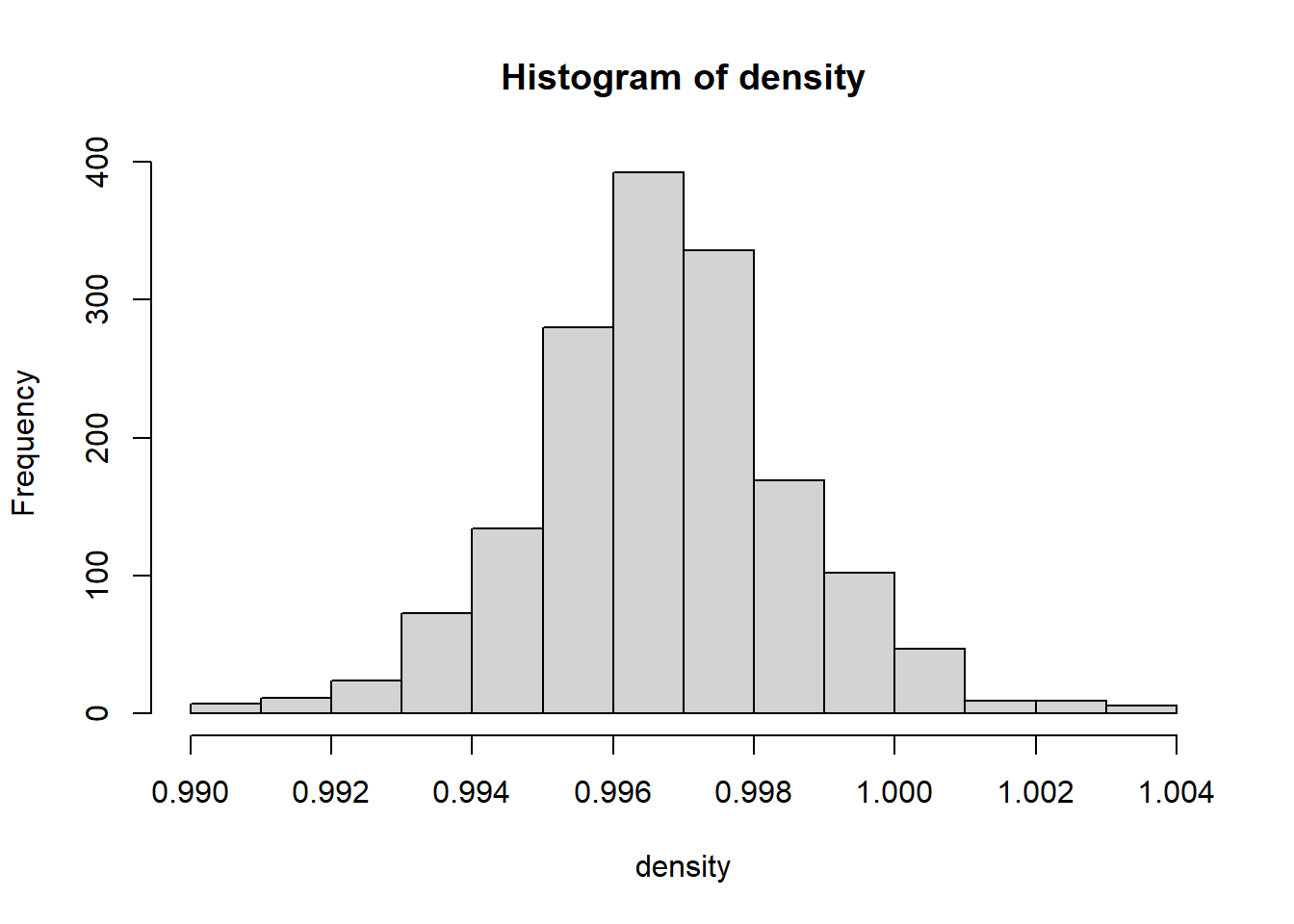
Note: To find the MLEs, you may consider multiplying the data by 10. This will solve a numerical problem you may encounter. You must transform back in the end.

wine<-read.csv(file="D:/Applied Stat methods/winequality-red.csv", header=T, sep = ";")

**attach**(wine)

**library**(stats4)

hist(density)



density\_10 <- density\*10

density\_10

###Last 100 Observations of density\_10.

# [1450] 9.9472 9.9470 9.9732 9.9374 9.9706 9.9974 9.9467 9.9236 9.9706

## [1459] 9.9603 9.9458 9.9724 9.9664 9.9545 9.9522 9.9580 9.9580 9.9728

## [1468] 9.9648 9.9728 9.9705 9.9578 9.9334 9.9656 9.9336 10.0242 9.9182

## [1477] 10.0242 9.9182 9.9808 9.9828 9.9498 9.9828 9.9719 9.9542 9.9592

## [1486] 9.9606 9.9546 9.9496 9.9420 9.9448 9.9344 9.9420 9.9348 9.9636

## [1495] 9.9459 9.9492 9.9636 9.9508 9.9582 9.9508 9.9508 9.9642 9.9638

## [1504] 9.9555 9.9605 9.9600 9.9562 9.9605 9.9814 9.9410 9.9661 9.9712

## [1513] 9.9588 9.9294 9.9842 9.9842 9.9633 9.9489 9.9647 9.9665 9.9489

## [1522] 9.9585 9.9633 9.9530 9.9522 9.9552 9.9553 9.9510 9.9714 9.9608

## [1531] 9.9444 9.9631 9.9573 9.9717 9.9397 9.9590 9.9528 9.9484 9.9538

## [1540] 9.9714 9.9646 9.9666 9.9472 9.9758 9.9550 9.9531 9.9575 9.9306

## [1549] 9.9783 9.9419 9.9783 9.9768 9.9586 9.9765 9.9627 9.9514 9.9636

## [1558] 9.9627 9.9787 9.9622 9.9622 9.9622 9.9546 9.9546 9.9546 9.9489

## [1567] 9.9494 9.9546 9.9629 9.9396 9.9340 9.9514 9.9632 9.9467 9.9677

## [1576] 9.9474 9.9588 9.9622 9.9540 9.9402 9.9470 9.9402 9.9362 9.9578

## [1585] 9.9484 9.9494 9.9492 9.9483 9.9414 9.9770 9.9314 9.9402 9.9574

## [1594] 9.9651 9.9490 9.9512 9.9574 9.9547 9.9549

sd\_density\_10 <- sd(density\_10)

dnorm(density\_10[1], mean=mean(density\_10), sd=sd(density\_10))

## [1] 18.08945

minuslog.lik<-**function**(mu, sigma){

log.lik<-0

**for**(i **in** 1: length(density\_10)) {

log.lik<-log.lik+log(dnorm(density\_10[i], mean=mu,

sd=sigma))

}

**return**(-log.lik)

}

minuslog.lik(80, 20)

## [1] 16062.54

minuslog.lik(100,30)

## [1] 14108.6

minuslog.lik(0,1)

## [1] 80900.31

est <- stats4::mle(minuslog = minuslog.lik,

start = list(mu = mean(density\_10),

sigma = sd(density\_10)),

lower=c(0, 0))

summary(est)

## Maximum likelihood estimation

##

## Call:

## stats4::mle(minuslogl = minuslog.lik, start = list(mu = mean(density\_10),

## sigma = sd(density\_10)), lower = c(0, 0))

##

## Coefficients:

## Estimate Std. Error

## mu 9.96746679 0.0004719810

## sigma 0.01887334 0.0003296881

##

## -2 log L: -8159.31

#Transform back

*### Transform back*

mle\_mean <- est@coef["mu"] / 10 *# Divide by 10 to undo scaling*

mle\_sd <- est@coef["sigma"] / 10 *# Divide by 10 to undo scaling*

mle\_mean

## mu

## 0.9967467

mle\_sd

## sigma

## 0.001887334

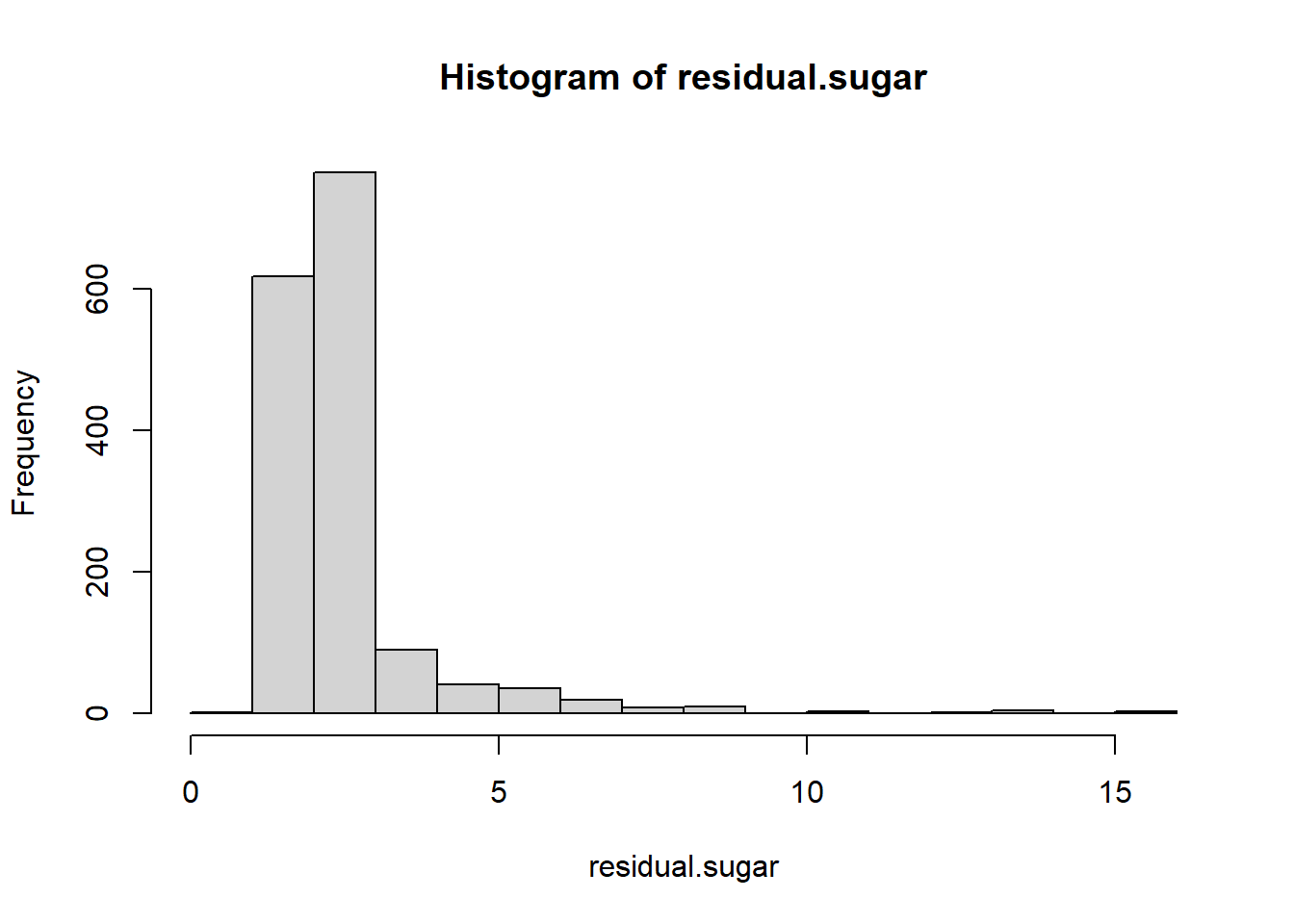
With an implication of density\*10, we get a new set of observations of 1599, I have copied the last 100 observations to keep it clear. Yes, we can use a normal distribution. The maximum likelihood estimates are 9.976 and 0.018 for the mean and standard deviation. The standard error for the mean is 0.000472 and 0.000330 for the standard deviation. We can also see the mean and standard deviation values after we transform the values to the original set.

2. Suppose the population mean of the variable “residual sugar” is μ , answer the following

questions.

1. Can we use a normal distribution to model “residual sugar”? If no, what distribution do you think can approximate its empirical distribution? What parameters are needed to characterize such a distribution? what are their maximum likelihood estimates? Please provide their standard errors as well?

hist(residual.sugar)



minuslog.lik <- **function**(mu, sigma) {

log.lik <- 0

**for**(i **in** 1:length(residual.sugar)) {

log.lik = log.lik + log(dlnorm(x=`residual.sugar`[i], meanlog=mu, sdlog=sigma))

}

**return**(-log.lik)

}

estimate\_sugar <- mle(minuslog=minuslog.lik,

start = list(mu=log(mean(`residual.sugar`)), sigma=log(sd(`residual.sugar`))),lower=c(0,0))

summary(estimate\_sugar)

## Maximum likelihood estimation

##

## Call:

## mle(minuslogl = minuslog.lik, start = list(mu = log(mean(residual.sugar)),

## sigma = log(sd(residual.sugar))), lower = c(0, 0))

##

## Coefficients:

## Estimate Std. Error

## mu 0.8502316 0.008936108

## sigma 0.3573326 0.006318587

##

## -2 log L: 3965.773

From the histogram, we can see that the distribution is skewed to the right. With the use of log functions, the maximum likelihood estimates are observed at 0.852316 and 0.3573326 for the mean and standard deviation. The standard error for the mean and standard deviation are observed at 0.008936108 and 0.006318587. With this data, we can say that we cannot use a normal distribution model for residual sugar.

3. We classify those wines as “excellent” if their rating is at least 7. Suppose the population the proportion of excellent wines is p.

1. What is the maximum likelihood estimate of p and its standard error?

Note: you need to create a new column of data for the new variable “excellent” which has binary values (0 and 1). Its value is 1 if the wine rating >=7, and otherwise its value is 0. Then, you can make inference about the underlying proportion parameter p using this column of data.

minuslog.lik<-**function**(p){

log.lik<-0

**for**(i **in** 1:length(wine$excellent)) {

log.lik<-log.lik+log(dbinom(wine$excellent[i], size = 1, prob = p))

}

**return**(-log.lik)

}

est <- stats4::mle(minuslog = minuslog.lik, start = list(p = hat))

summary(est)

## Maximum likelihood estimation

##

## Call:

## stats4::mle(minuslogl = minuslog.lik, start = list(p = hat))

##

## Coefficients:

## Estimate Std. Error

## p 0.1357098 0.008564278

##

## -2 log L: 1269.921

The maximum likelihood estimation using the log function is observed as 0.1357098 for mean and 0.008564278 for standard deviation for function p.