Wine_Project_Part B_2

Sohail Shaikh

2023-10-01

```
library(stats4)
library(here)
library(dplyr)
```

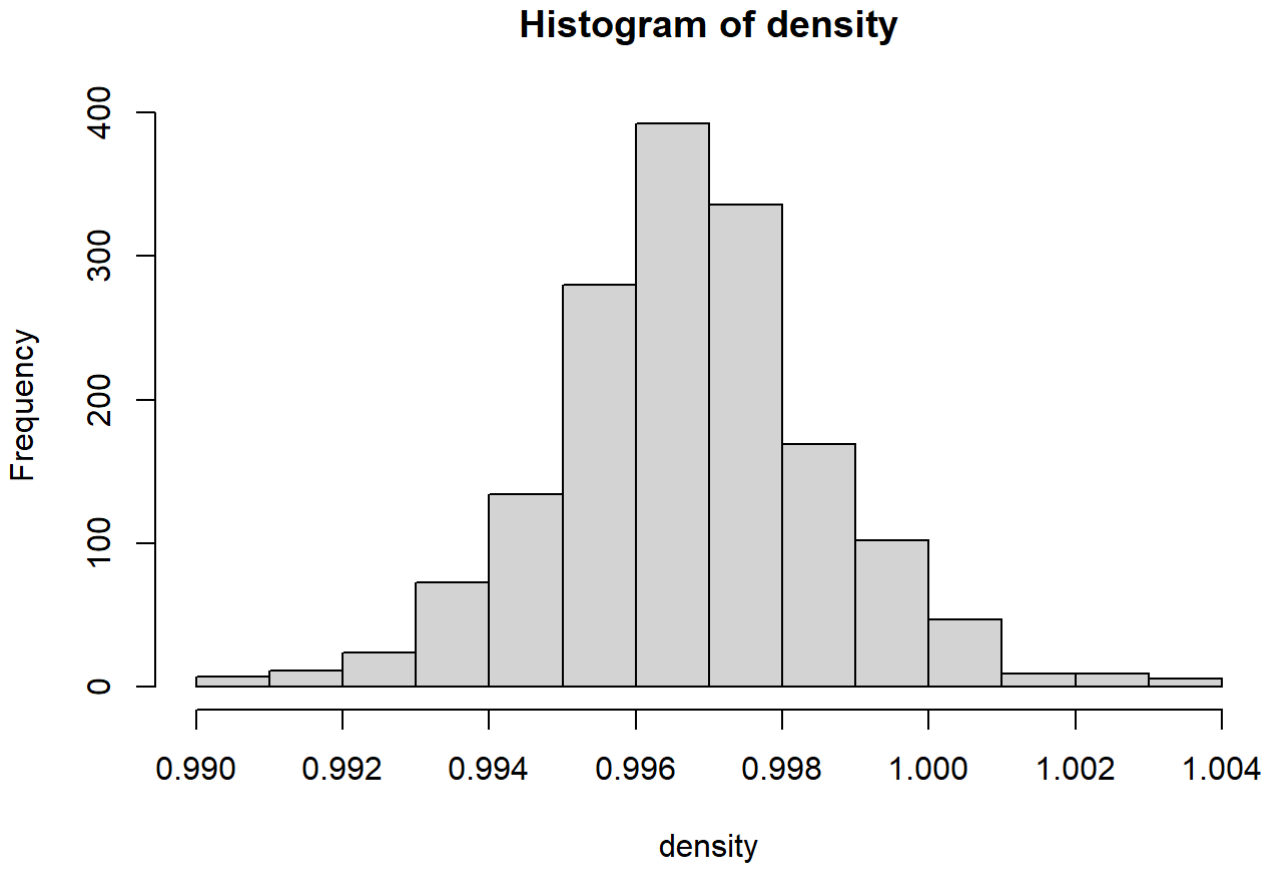
Loading Data

```
#importing wine data


df <- read.csv(file = "E:\\Linder_college\\Data Wrangling\\R_scripts\\data\\winequality-red.csv", sep = ";",header = T)

attach(df)
```

Question 1e

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.

```
hist(density)
```



Looking at the above plot we can say that the density variable is fairly normally distributed.

Calculating the maximum likelihood estimates of the mean and standard deviation for density

```
density_scl <- density * 10

minuslog.lik <- function(mu, sigma) {
  log.lik <- 0
  for(i in 1:1599) {
    log.lik <- log.lik + log(dnorm(density_scl[i], mean = mu, sd = sigma))
  }
  return(-log.lik)
}

estimate_1 <- mle(minuslog = minuslog.lik,
           start = list(mu = mean(density_scl),
           sigma = sd(density_scl)))

estimated_params <- coef(estimate_1)
estimated_params["mu"] <- estimated_params["mu"] / 10
estimated_params["sigma"] <- estimated_params["sigma"] / 10

summary(estimate_1)
```

```
## Maximum likelihood estimation
##
## Call:
## mle(minuslogl = minuslog.lik, start = list(mu = mean(density_scl),
##     sigma = sd(density_scl)))
##
## Coefficients:
##         Estimate   Std. Error
## mu    9.96746679 0.0004719810
## sigma 0.01887334 0.0003296881
##
## -2 log L: -8159.31
```

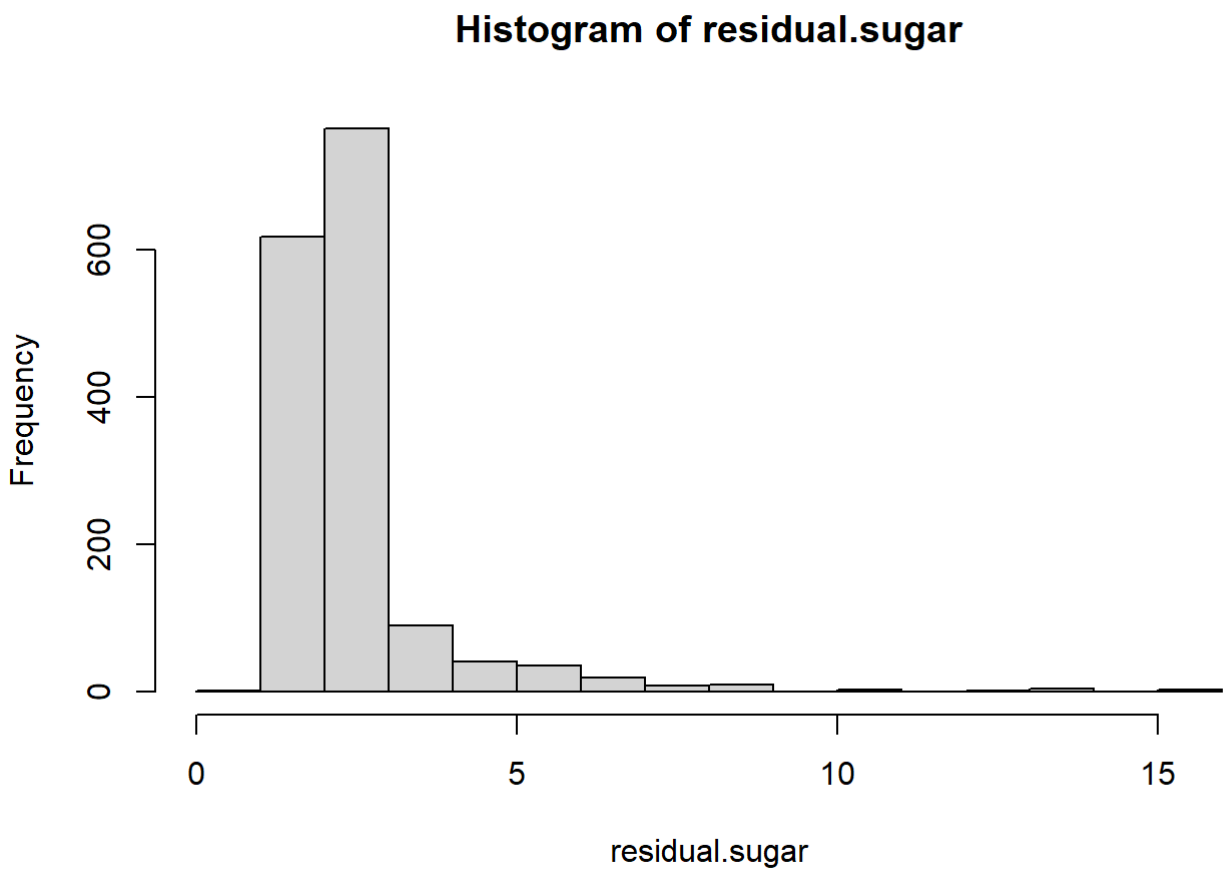
Maximum likelihood estimate of mean for density is 9.96746679 and its standard error is 0.0004719810

Maximum likelihood estimate of standard deviation for density is 0.01887334 and its standard error is 0.0003296881

Question 2d

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)
```



Looking at the plot we can say that the distribution is highly skewed, i think lognormal distribution can approximate its empirical distribution.

```
minuslog.lik1 <- function(mu, sigma) {
  log.lik <- 0
  for(i in 1:1599) {
    log.lik <- log.lik + log(dlnorm(`residual.sugar`[i], meanlog = mu, sdlog = sigma))
  }
  return(-log.lik)
}

estimate_2 <- mle(minuslog = minuslog.lik1,
           start = list(mu = log(mean(`residual.sugar`)),
           sigma = log(sd(`residual.sugar`))))

summary(estimate_2)
```

```
## Maximum likelihood estimation
##
## Call:
## mle(minuslogl = minuslog.lik1, start = list(mu = log(mean(residual.sugar)),
##     sigma = log(sd(residual.sugar))))
##
## Coefficients:
##        Estimate  Std. Error
## mu    0.8502341 0.008935942
## sigma 0.3573260 0.006318293
##
## -2 log L: 3965.773
```

Maximum likelihood estimate of mean for residual sugar is 0.8502341 and its standard error is 0.008935942

Maximum likelihood estimate of standard deviation for residual sugar is 0.3573260 and its standard error is 0.006318293

Question 3d

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

```
p_df <- df %>%
  mutate(excellent = ifelse(quality >= 7, 1, 0))

excellent <- p_df$excellent
```

```
minuslog.lik2 <- function(p){ #likelihood function
  log.lik <- 0
  for (i in 1:1599) {
    log.lik <- log.lik + log(dbinom(excellent[i], size = 1, prob = p))
  }
  return(-log.lik) #dbinom for binomial distribution
  #taking log of function
}

#taking mle
estimate.rbinom <- mle(minuslog = minuslog.lik2, start = list(p = mean(excellent)))
summary(estimate.rbinom)
```

```
## Maximum likelihood estimation
##
## Call:
## mle(minuslogl = minuslog.lik2, start = list(p = mean(excellent)))
##
## Coefficients:
##    Estimate  Std. Error
## p 0.1357098 0.008564278
##
## -2 log L: 1269.921
```

Maximum likelihood estimate for p is 0.1357098 and its standard error is 0.008564278