Q2_MidTerm_Aditya_Mhaske.R

adityamhaske

2023-03-09

Aditya Sanjay Mhaske

Questiion 2

#'

A) Calculate the plug-in estimates of the mean, median, variance, and interquartile range.

```
library(ggplot2)

data(ChickWeight)

set.seed(100)
x <- sample(ChickWeight$weight, size = 60, replace = TRUE)
```

plugin estimates

```
n=length(x)
n
```

```
## [1] 60
```

```
fx = rep(1,n) / n
fx
```

```
##  [1] 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667
##  [7] 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667
## [13] 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667
## [19] 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667
## [25] 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667
## [31] 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667
## [37] 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667
## [43] 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667
## [49] 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667
## [55] 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667 0.01666667
```

```
EX = sum(x * fx)
EX
```

```
## [1] 115.0833
```

```r
mean(x)
```

```
## [1] 115.0833
```

```r
median(x)
```

```
## [1] 113.5
```

```r
VarX = sum ((x - EX)^2 * fx)
VarX
```

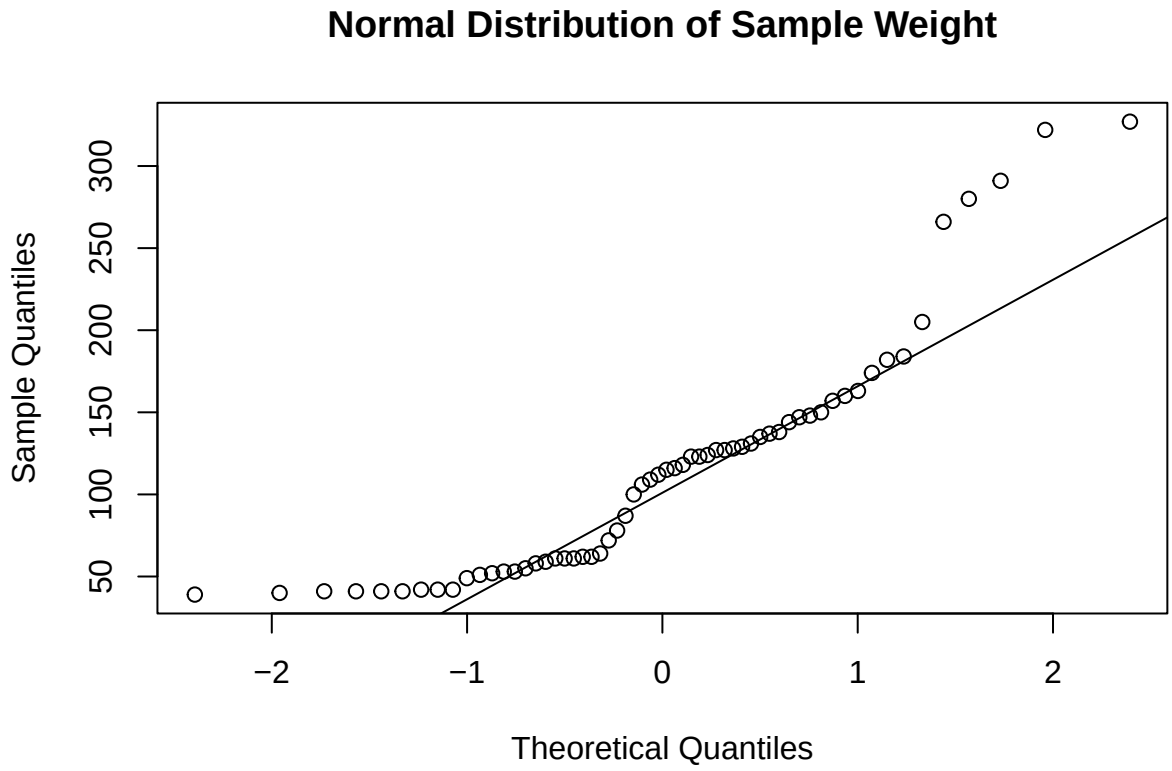
```
## [1] 5052.41
```

```r
IQR(x)
```

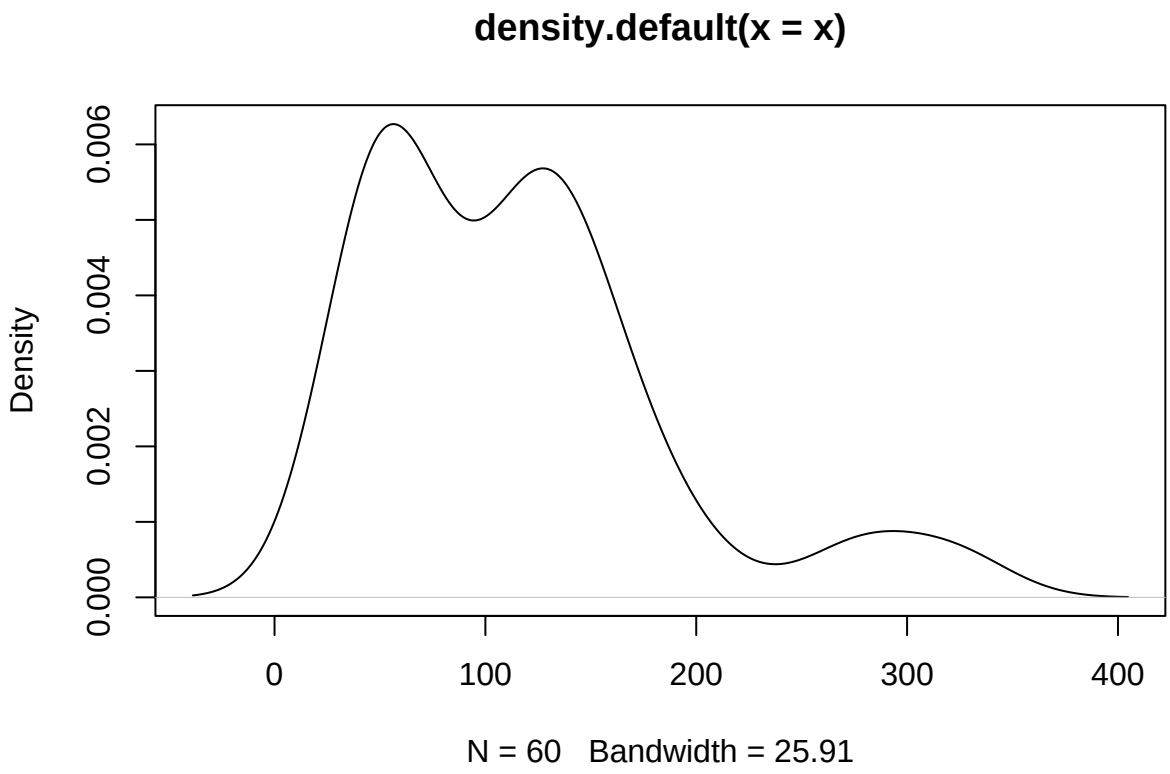
```
## [1] 87.5
```

B) Do you think that the sample was drawn from a normal distribution? Justify your answer.

```r
qqnorm(x, main = "Normal Distribution of Sample Weight")
qqline(x)
```



```
plot(density(x))
```



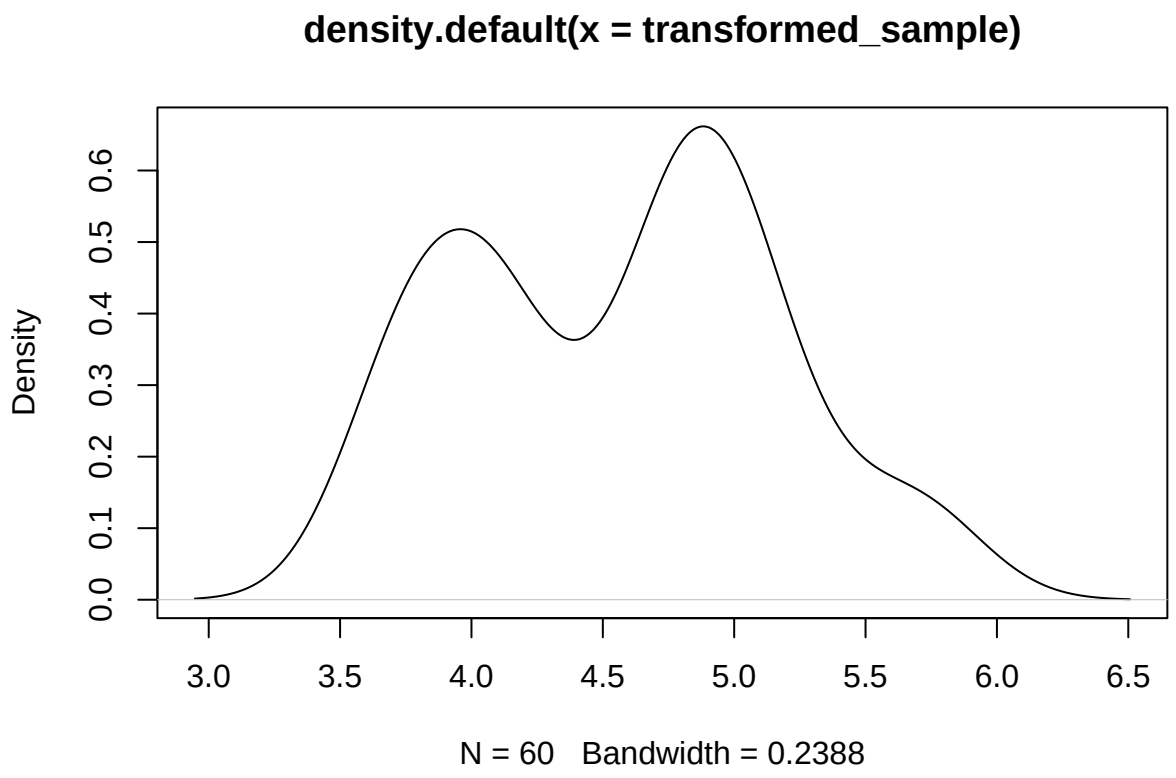
Based on the qqplot and the kernel density plot, we can see that the sample isn't drawn from a normal distribution

---

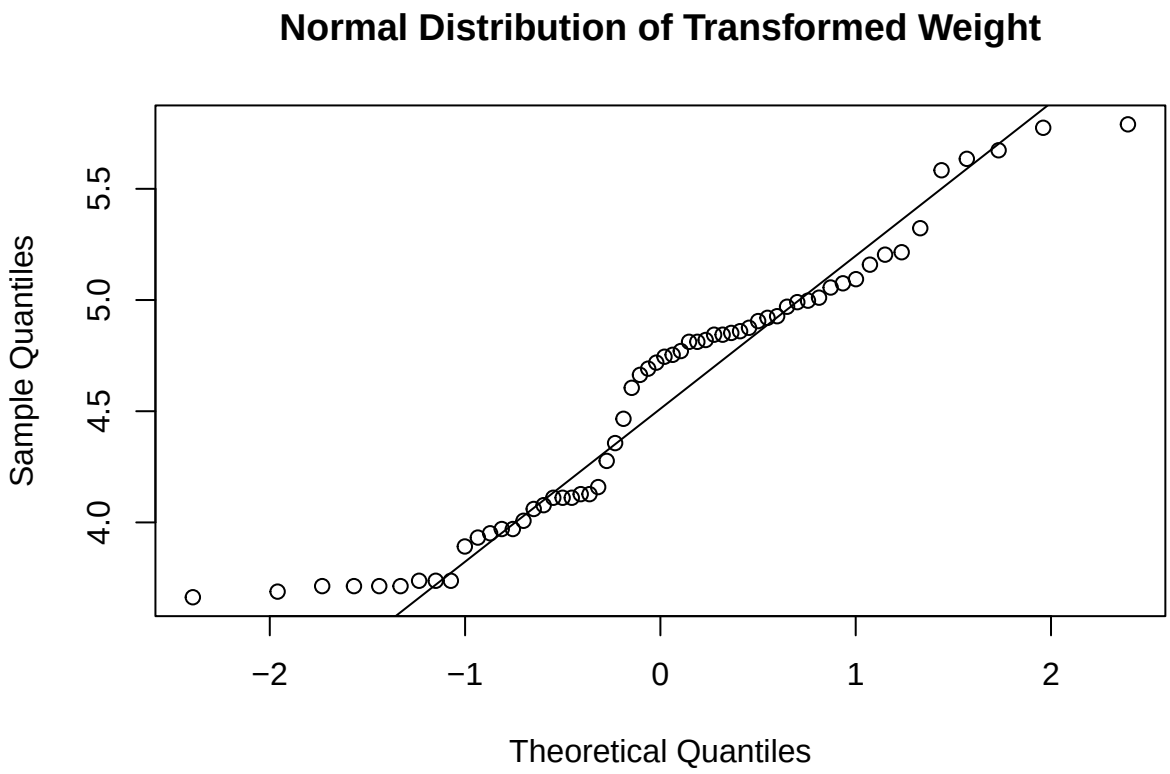
C) Now consider the transformed sample produced by replacing each value with its natural logarithm. Do you think that the transformed sample was drawn from a normal distribution? Justify your answer.

```
# Transformed Weight
transformed_sample <- log(x)

plot(density(transformed_sample)) #Kernel Density Plot
```




```
#qqplot
qqnorm(transformed_sample, main = "Normal Distribution of Transformed Weight")
qqline(transformed_sample)
```



From the Kernel Density Plot and the qqplot, we can illustrate that the transformed sample weight distribution is more symmetric than part (C), this is why we can say that the transformed sample was drawn from a normal distribution.

D. Calculate the test statistic

```
x_bar <- 116
s <- 70
n <- 60
se <- s/sqrt(n)
t <- (x - 150)/se
t
```

```
##  [1]  -2.5451033  -2.3237900   3.7623267   2.6557600 -10.0697567  -4.8688933
##  [7] -11.9509200  -3.8729833  -1.3278800  -2.8770733 -10.1804134   3.5410133
## [13] -10.7336967 -10.8443534  14.3853667   0.0000000  -0.2213133  -7.9672800
## [19]  -2.4344467  -0.3319700 -11.9509200  -2.9877300  12.8361734 -12.0615767
## [25] -10.5123834  -0.6639400  -9.7377867  -9.8484434  -9.7377867  -2.5451033
## [31] -12.2828900 -12.0615767  -2.1024767  -6.9713700  -3.5410133  -9.5164734
## [37] -12.1722334 -12.0615767  -2.9877300  15.6025901  19.5862301  -4.2049533
## [43]  -1.4385367  -4.5369233   1.4385367   1.1065667  -9.8484434   6.0861167
## [49] -10.9550100  -8.6312200  -3.7623267 -11.1763234 -10.7336967 -12.0615767
## [55] -11.9509200  19.0329467  -9.8484434  -5.5328334  -1.6598500   0.7745967
```

The value of the test statistic is -2.595893, indicating the number of standard errors by which x differs from the assumed mean of 150. To determine the p-value for this test statistic, we can utilize the pt() function available in R.

```
p_value <- pt(t, df = n-1, lower.tail = TRUE)
p_value
```

```
##  [1] 6.779727e-03 1.180053e-02 9.998053e-01 9.949201e-01 9.903330e-15
##  [6] 4.380406e-06 1.069647e-17 1.359640e-04 9.466616e-02 2.789620e-03
## [11] 6.552248e-15 9.996072e-01 8.475636e-16 5.653801e-16 1.000000e+00
## [16] 5.000000e-01 4.128063e-01 3.107703e-11 8.979516e-03 3.705442e-01
## [21] 1.069647e-17 2.045306e-03 1.000000e+00 7.253654e-18 1.912817e-15
## [26] 2.546575e-01 3.445506e-14 2.271041e-14 3.445506e-14 6.779727e-03
## [31] 3.351285e-18 7.253654e-18 1.989408e-02 1.512311e-09 3.927818e-04
## [36] 7.958925e-14 4.926590e-18 7.253654e-18 2.045306e-03 1.000000e+00
## [41] 1.000000e+00 4.495172e-05 7.778283e-02 1.427687e-05 9.222172e-01
## [46] 8.635136e-01 2.271041e-14 1.000000e+00 3.776864e-16 2.363033e-12
## [51] 1.946894e-04 1.692809e-16 8.475636e-16 7.253654e-18 1.069647e-17
## [56] 1.000000e+00 2.271041e-14 3.796396e-07 5.112561e-02 7.791644e-01
```

The p-value is 0.005772819 As the obtained p-value is below the typical significance threshold of 0.05, we have sufficient evidence to reject the null hypothesis and understand that the mean weight is possibly less than 150 grams. Therefore, the claim that the mean weight of the chicks in the entire population is no less than 150 grams is not supported by the available data.
