Homework 10 - STAT440

Joseph Sepich (jps6444)

11/15/2020

```
set.seed(42)
```

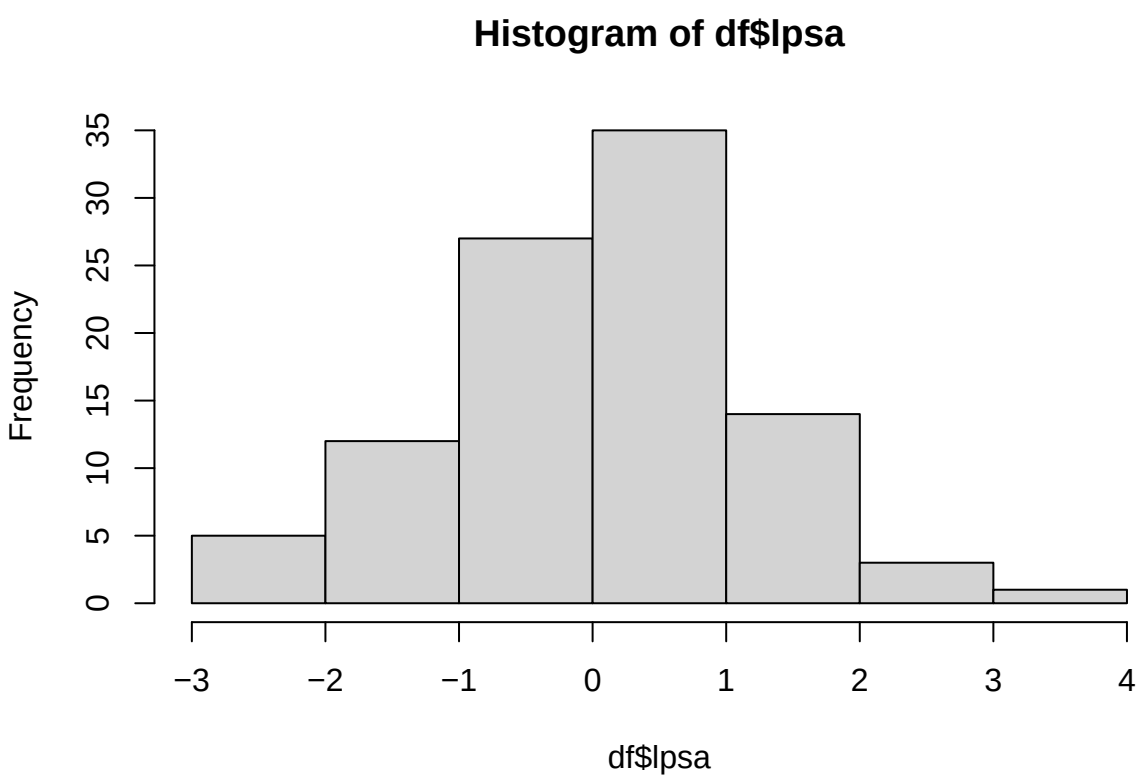
```
df <- read.csv('./data/prostate.csv')
head(df)
```

```
##   X     lcavol    lweight        age      lbph        lcp    gleason      lpsa
## 1 1 -1.6373556 -2.0062118 -1.8624260 -1.024706 -0.8631712 -1.0421573 -2.909170
## 2 2 -1.9889805 -0.7220088 -0.7878962 -1.024706 -0.8631712 -1.0421573 -2.640906
## 3 3 -1.5788189 -2.1887840  1.3611634 -1.024706 -0.8631712  0.3426271 -2.640906
## 4 4 -2.1669171 -0.8079939 -0.7878962 -1.024706 -0.8631712 -1.0421573 -2.640906
## 5 5 -0.5078745 -0.4588340 -0.2506313 -1.024706 -0.8631712 -1.0421573 -2.106823
## 6 6 -2.0361285 -0.9339546 -1.8624260 -1.024706 -0.8631712 -1.0421573 -1.712919
```

Problem 1

Part a

```
hist(df$lpsa)
```



I am going to use the normal distribution as the form of our liklihood function. The data is symmetric and has a single mode in the center. This mimics the form of a normal distribution.

Part b

Note that we will use the sample variance as our "known", fixed variance value.

We will use a normal distribution as our prior distribution for our mean. We will use this as it is the conjugate prior for the normal distribution likelihood, so the posterior will also be a normal distribution.

Part c

A large portion of the values are near 0, so we will use the prior parameter $\mu_0 = 0$. It is likely that the 0 level is considered the "normal" level for measurement purposes as (via very brief research) there is a certain level that doctors use as a cutoff for whether further tests are needed (in the non-log variable). The sample mean is also very close to this value.

```
mean(df$lpsa)
```

```
## [1] -1.969206e-15
```

Part d

Part e

Problem 2

Part a

Part b

Part c

Part d

Part e