BSDA: Assignment 3

September 16, 2023

```
knitr::opts_chunk$set(echo = TRUE)
```

Times used for reading and self-study exercise: 8
Time used for the assignment: 17
Good with the assignment: Good that the new functions for t-distribution was included in the package.
Things to improve in the assignment: I did not found the Frank Harrell recommendation so helpful

```
library("bsda")
data("windshieldy1")
head(windshieldy1)
```

```
## [1] 13.357 14.928 14.896 15.297 14.820 12.067
```

```
windshieldy_test <- c(13.357, 14.928, 14.896, 14.820)
```

Question 1: Inference for nomal mean and variance

Q1a, Bayes point estimate $E(\mu|y)$ and 95 % posterior interval

For this subtask we are interested in inference for parameter μ so for now parameter element σ is considered a nuisance parameter. First, we will compute the Bayes point estimate for parameter μ, ie $E(\mu|y)$ and secondly calculate the 95% posterior interval. Lastly, we plot the density function for the posterior distribution

We have two unknown parameter element in the parameter vector which gives us $\boldsymbol{\theta} = (\mu, \sigma^2)$. Below we state the prior distribution, the likelihood function and the resulting posterior distribution:

Noninformative prior distribution:
$p(\boldsymbol{\theta}) = p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$

Likelihood function:
$L(y|\boldsymbol{\theta}) = L(y|\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}\left(y_i - \mu\right)^2\right)$

The Joint posterior distribution:
$P(\boldsymbol{\theta}|y) = P(\mu, \sigma^2|y) \propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y} - \mu)^2\right]\right)$,
where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$

1

Worth mentioning is that s^2 is an unbiased estimator of parameter σ^2 when parameter μ is unknown.

The Marginal posterior distribution for parameter μ is being calculated according to below:

$$p(\mu \mid y) = \int_0^\infty p\left(\mu, \sigma^2 \mid y\right) d\sigma^2 \propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2}\right]^{-n/2}$$

$$= t_{n-1}\left(\mu | \bar{y}, \frac{s}{\sqrt{n}}\right)$$

Now when the posterior, prior and likelihood has been defined we create a function that computes the point estimate for parameter μ, ie $E(\mu|y)$:

```
mu_point_est <- function(data = windshieldy_test) {

  # calculating the expected value for the (mu | y)-distribution.
  pointest <- round(mean(data), digits = 2)
  return(pointest)


}
mu_point_est(data = windshieldy1)

## [1] 14.61
```

So the expected value of parameter μ given our data y corresponds to $E(\mu|y) = 14.61$ and the function below computes the 95 % posterior interval.

```
mu_interval <- function(data = windshieldy_test, prob = 0.95) {

  n <- length(data)

  sd <- sqrt(var(data) / length(data))

  mean <-  mean(data)

  output <- round(qt(c( (1-prob)/2, 1-(1-prob)/2 ),
             df=n-1 ) * sd + mean, digits = 1)
  return(output)
}
mu_interval(data = windshieldy1)

## [1] 13.5 15.7
```

and the 95 % posterior interval is $[13.5, 15.7]$ such that $P(13.5 < \mu < 15.7) \approx 0.95$. The density plot for the posterior looks like follows:

```
# plotting the density function for the t-distribution
dataset <- windshieldy1
n <- length(dataset)
sd <- sqrt(var(dataset) / length(dataset))
mean <-  mean(dataset)

parameter_values <- seq(0,1, by = 0.001)
quantiles <- qtnew(p = parameter_values,
               df = n - 1,
```
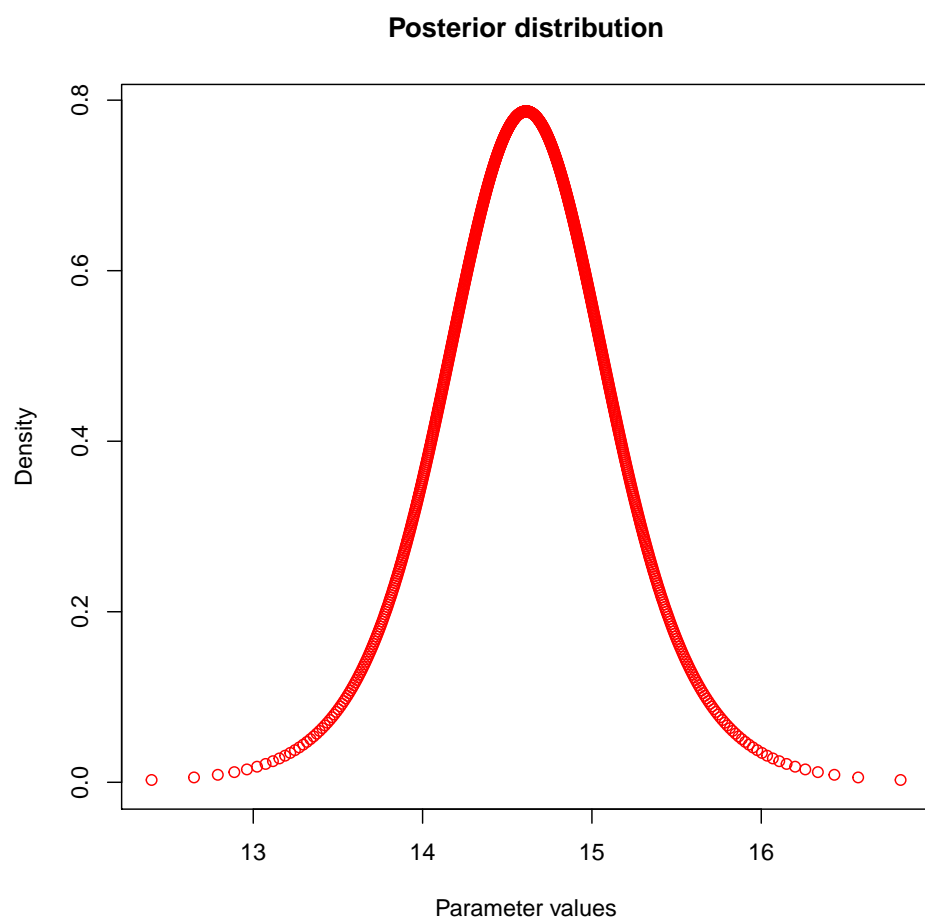
```
                          mean = mean,
                          scale = sd)

post_dist <- dtnew(x = quantiles, df = n-1, mean = mean, scale = sd)

plot(x = quantiles,
     y = post_dist,
     col = "red",
     main = "Posterior distribution",
     ylab = "Density",
     xlab = "Parameter values")
```

Posterior distribution

Q1b, Bayesian point estimate, 95 % predictive interval and plot the density

We start by computing the point estimate for $\tilde{y}$, i.e the $E(\tilde{y}|y)$ and the posterior predictive distribution when having unknown mean and variance is $p(\tilde{y} \mid y) = t_{n-1}\left(\tilde{y} \mid \bar{y}, \left(1 + \frac{1}{n}\right)s^2\right)$. Hence, we got that

$E(\tilde{y}|y) = \bar{y}$ and
$Var(\tilde{y}|y) = \left(1 + \frac{1}{n}\right) \cdot s^2$

The function below computes the posterior predictive point estimate

```
# function for posterior predictive point estimate
mu_pred_point_est <- function(data  = windshieldy_test) {

  output <- round(mean(data), digits = 2)
  return(output)
}
mu_pred_point_est(data = windshieldy1)

## [1] 14.61
```

The posterior predictive point estimate corresponds to $E(\tilde{y}|y) = 14.6$ and now we create a function that computes the 95 % predictive interval.

```
# function for predictive posterior interval
mu_pred_interval <- function(data = windshieldy_test,
                             prob = 0.95) {

  n <- length(data)

  sd <- sqrt( (1+ 1/n)*var(data))

  mean <-  mean(data)

  output <- round(qt(c( (1-prob)/2, 1-(1-prob)/2 ),
             df=n-1 ) * sd + mean,
             digits = 1)

  return(output)

}
mu_pred_interval(data = windshieldy1)

## [1] 11.0 18.2
```

and the 95 % predictive interval is $[11.0, 18.2]$ such that $P(11.0 < \mu < 18.2) \approx 0.95$. Worth mentioning is that a predictive interval is different from a posterior interval. Lastly for this subtask, we plot the density for the resulting posterior predictive distribution.

```
n <- length(windshieldy1)
mean <- mean(windshieldy1)
sd <- sqrt( (1+ 1/n)*var(windshieldy1))
```

4

```
parameter_values <- seq(0,1, by = 0.001)
quantiles_values <- qtnew(p = parameter_values,
                          df = n-1,
                          mean = mean,
                          scale = sd)

posterior_pred_distribution <- dtnew(x = quantiles_values,
                                     df = n-1,
                                     mean = mean,
                                     scale = sd)

plot(x = quantiles_values,
     y = posterior_pred_distribution,
     col = "blue",
     main = "Posterior predictive distribution",
     xlab = "Predictive values",
     ylab = "Density")
```
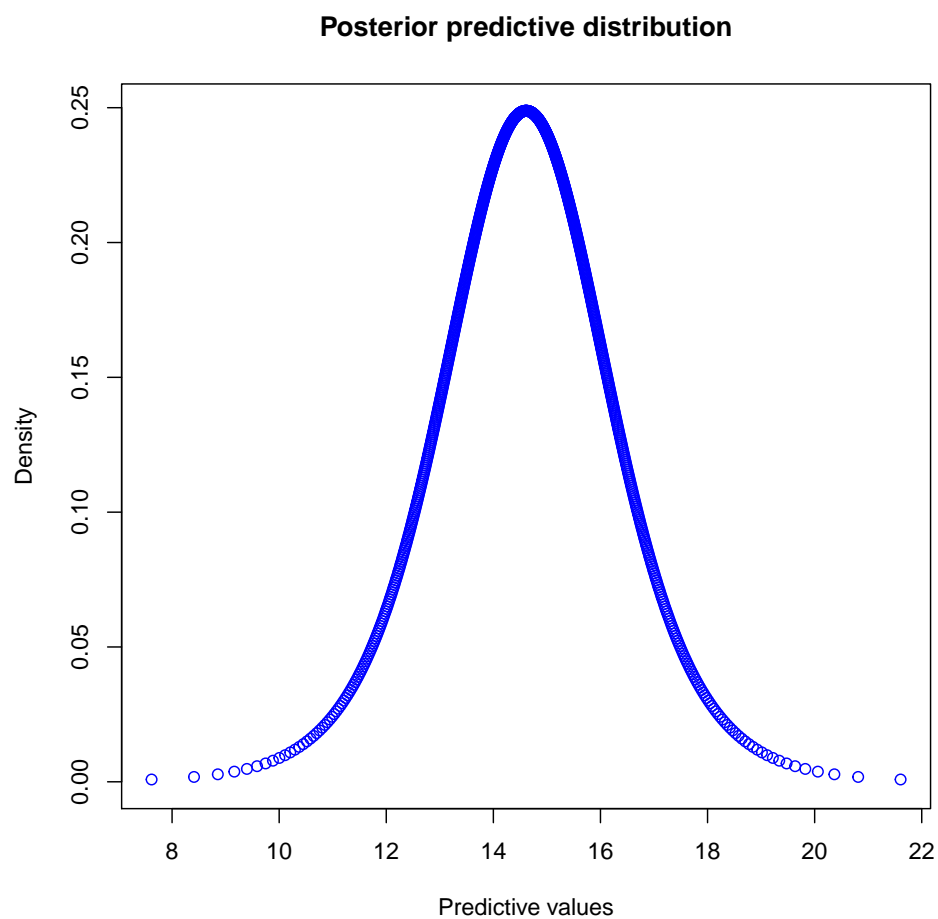

Posterior predictive distribution

Question 2

Q2a, Inference for the difference between proportions

For this subtask we will compare the difference between the proportion for treatment group and the proportion for control group. The outcomes are independent and binomially distributed. So we have that $y_i \mid \theta_i \sim \text{Bin}(n_i, \theta_i)$.

Hence, due to independent between groups and groups of k = 2 we have the following likelihood function, noninformative prior distribution and resulting posterior distribution:
Likelihood function for both groups:
$L(y_i|p_i, n_i) = \prod_{i=0}^{1} p(y_i \mid p_i, n_i) = p(y_0 \mid p_0, n_0) \cdot p(y_1 \mid p_1, n_1)$
$\propto p_0^{y_0} \cdot (1 - p_0)^{n_0 - y_0} \cdot p_1^{y_1} \cdot (1 - p_1)^{n_1 - y_1}$

The noninformative prior distribution for p_0 and p_1:
$p(p_0, p_1) \propto 1 \sim Unif(0, 1)$

The resulting posterior distribution for p_0 and p_1:
$p(p_0, p_1|y_i, n_i) \propto p(p_0, p_1) \cdot \prod_{i=0}^{1} p(y_i \mid p_i, n_i) \propto p_0^{y_0} \cdot (1 - p_0)^{n_0 - y_0} \cdot p_1^{y_1} \cdot (1 - p_1)^{n_1 - y_1}$

The marginal posterior distribution for p_0:

$p(p_0|y_0, n_0) = p_0^{y_0} \cdot (1 - p_0)^{n_0 - y_0} \underbrace{\int_0^1 p_1^{y_1} \cdot (1 - p_1)^{n_1 - y_1} \, dp_1}_{p_1 \sim Beta(y_1+1, n_1-y_1+1)}$

$p_0^{y_0} \cdot (1 - p_0)^{n_0 - y_0} \sim Beta(y_0 + 1, n_0 - y_0 + 1)$

Because $\boldsymbol{Y_0}$ and $\boldsymbol{Y_1}$ are independent to each other and their joint distribution can be written as a product of respective probability mass functions, ie $P(y_0, y_1) = P(y_0) \cdot P(y_1)$ then the same process can be done to find the marginal posterior distribution for p_1. But instead of integrating with respect to p_1 we integrate with respect to p_0 to find the marginal posterior distribution for p_1 which gives us the following

The marginal posterior distribution for p_1:
$p(p_1|y_1, n_1) \sim Beta(y_1 + 1, n_1 - y_1 + 1)$

Now when the likelihood function, the noninformative prior and both joint and marginal resulting posterior distributions has been stated we can summarize the result for the odds ratio by computing the point estimate, posterior interval and the a histogram. We start by computing the point estimate for the odds ratio, ie $(p_1 / (1 - p_1)) / (p_0 / (1 - p_0))$ according to below:

```
# point estimate for the odds ratio
set.seed(4711)
p0 <- rbeta(100000, shape1 = 39 + 1, shape2 = 674 - 39 + 1)
p1 <- rbeta(100000, shape1 = 22 + 1, shape2 = 680 - 22 + 1)

posterior_odds_ratio_point_est <- function(p0 = p0, p1 = p1) {

  output <- mean((p1/(1-p1)) / (p0/(1-p0))  )
```

```
    return(output)
}
posterior_odds_ratio_point_est(p0, p1)
```

```
## [1] 0.570978
```

So the point estimate for the odds ratio corresponds to 0.570978 Now we calculate the 95 % posterior interval for the odds ratio

```
# posterior interval

posterior_odds_ratio_interval <- function(p0 = p0,
                                          p1 = p1,
                                          prob = 0.90) {

  odds_ratio <- (p1/(1-p1)) / (p0/(1-p0))

  output <- round(quantile(odds_ratio,
                  probs = c( (1-prob)/2,
                            1-(1-prob)/2)), digits = 3)

  return(output)

}
posterior_odds_ratio_interval(p0, p1, prob = 0.95)
```

```
##  2.5% 97.5%
## 0.321 0.925
```
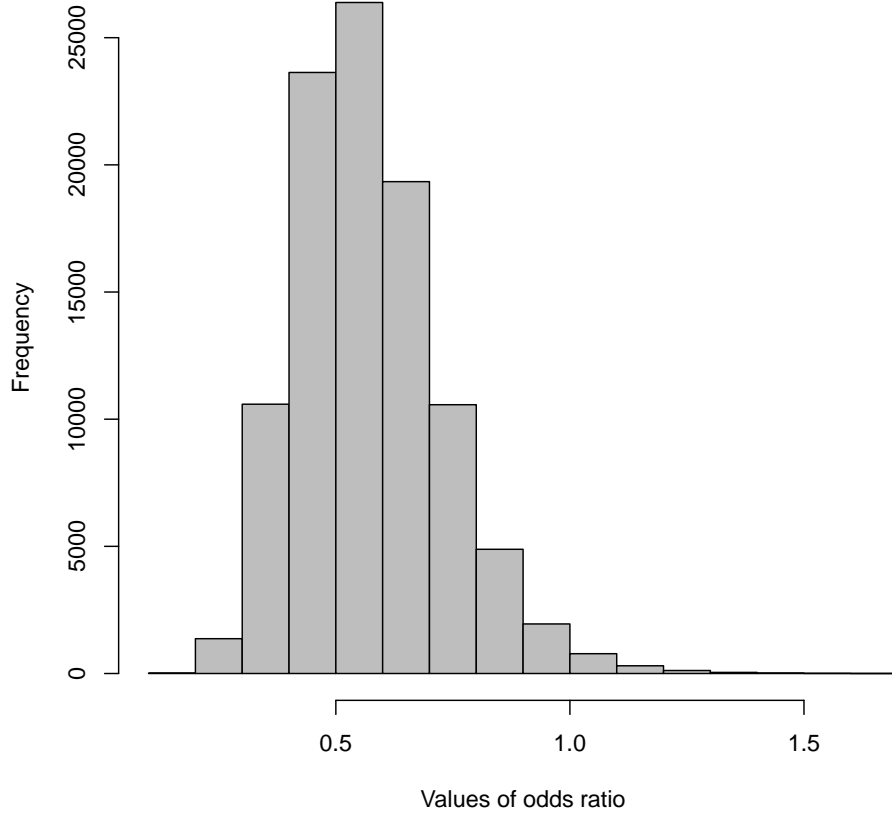
and the 95 % posterior interval is $[0.321, 0.925]$ such that $P(0.321 < (p_1/(1-p_1)) / (p_0/(1-p_0)) < 0.925) \approx 0.95$. Lastly for this subtask, we plot a histogram for the odds ratio.

```
# histogram for the posterior distribution for odds ratio
odds_ratio <- (p1/(1-p1)) / (p0/(1-p0))
hist(odds_ratio,
     main = "Histogram for the odds ratio",
     col = "gray",
     xlab = "Values of odds ratio")
```

Histogram for the odds ratio

Q2b, Sensitivity of the inference due to the choice of prior

For Q2b we will analyze the sensitivity of the inference due to the choice of the prior by using another prior distribution. Earlier we used $p(p_0, p_1) \propto 1 \sim Unif(0,1)$, which corresponds to $\beta(1,1)$, and calculated the posterior distribution. Now we will use Jeffreys prior distribution for $\boldsymbol{\theta} = (p_0, p_1)$ when Y_0 and Y_1 are _independent_ to each other.

Jeffreys prior distribution for $\boldsymbol{\theta} = (p_0, p_1)$:
$$p(\boldsymbol{\theta}) = p(p_0, p_1) \propto \underbrace{p_0^{-1/2}(1-p_0)^{-1/2}}_{p_0 \sim Beta(1/2,1/2)} \cdot \underbrace{p_1^{-1/2}(1-p_1)^{-1/2}}_{p_1 \sim Beta(1/2,1/2)}$$

Joint posterior distribution for $\boldsymbol{\theta} = (p_0, p_1)$ given $\boldsymbol{Y} = (Y_0, Y_1)$:
$$p(\boldsymbol{\theta}|\boldsymbol{Y}) = p(p_0, p_1|y_0, y_1) \propto \underbrace{p_0^{y_0+1/2-1}(1-p_0)^{n_0-y_0+1/2-1}}_{p_0 \sim Beta(y_0+1/2, n_0-y_0+1/2)} \cdot \underbrace{p_1^{y_1+1/2-1}(1-p_1)^{n_1-y_1+1/2-1}}_{p_1 \sim Beta(y_1+1/2, n_1-y_1+1/2)}$$

So the resulting marginal posterior distribution for p_0:
$$p_0 \sim Beta(y_0 + 1/2, n_0 - y_0 + 1/2)$$

and the resulting marginal posterior distribution for p_1:

$$p_1 \sim Beta(y_1 + 1/2, n_1 - y_1 + 1/2)$$

```r
# sensitivity analysis for posterior distribution and odds ratio
# draw samples from p0 and p1 distribution
set.seed(4711)
p0_jeff <- rbeta(100000, shape1 = 39 + 1/2, shape2 = 674 - 39 + 1/2)
p1_jeff <- rbeta(100000, shape1 = 22 + 1/2, shape2 = 680 - 22 + 1/2)
# calculating the point estimate for odds ratio
posterior_odds_ratio_point_est(p0 = p0_jeff, p1 = p1_jeff) # 0.5657736
```

```
## [1] 0.5657736
```

```r
# calculating the posterior interval for odds ratio
posterior_odds_ratio_interval(p0 = p0_jeff, p1 = p1_jeff, prob = 0.95)
```

```
##  2.5% 97.5%
## 0.317 0.924
```

```r
# Result: [0.317, 0.924]
# plotting a histogram for odds ratio
odds_ratio_jeff <- (p1_jeff/(1- p1_jeff)) / (p0_jeff/(1-p0_jeff))
hist(odds_ratio_jeff,
     main = "Histogram for odds ratio w. Jeffreys prior",
     col = "dark blue",
     xlab = "Values of odds ratio")
```

Histogram for odds ratio w. Jeffreys prior
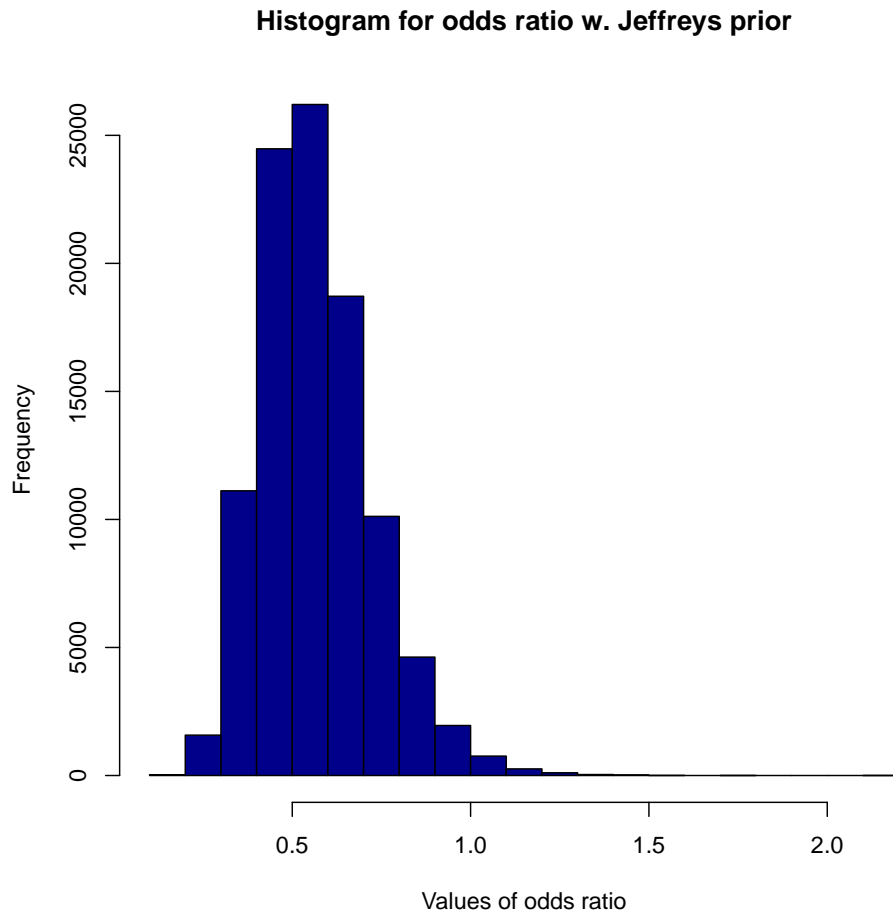
The histogram, point estimate and posterior interval for the odds ratio, which is based on the marginal posterior distribution for p_0 and p_1, when having Jeffreys prior is really similar to when having $unif(0,1)$ as the prior distribution. Hence, we are comparing the odds ratio when using two different noninformative priors, i.e priors that only uses general information about p_0 and p_1. Hence, we let the data speak for itself and the inference are unaffected by the information outside of the sample and therefore the odds ratio is very similar for different noninformative priors for a given sample, ie given $\boldsymbol{Y_0}$ and $\boldsymbol{Y_1}$.

Question 3, Inference for the difference b/w normal means

Q3a, Point estimate, Interval and plot for μ_d

We have been given two different datasets, named $\boldsymbol{Y_1}$ and $\boldsymbol{Y_2}$

```
data("windshieldy1") # dataset for y1 vector
data("windshieldy2") # dataset for y2 vector
```

and we assume independence between $\boldsymbol{Y_1}$ and $\boldsymbol{Y_2}$, ie indepedence between the two production lines. $\boldsymbol{Y_1}$ and $\boldsymbol{Y_2}$ got the following distributions:

$\boldsymbol{Y_1} \sim N(\mu_1, \sigma_1^2)$
$\boldsymbol{Y_2} \sim N(\mu_2, \sigma_2^2)$

We want to look at μ_d where $\mu_d = \mu_1 - \mu_2$ so we start by defining the likelihood, prior and posterior for $\boldsymbol{Y_1}$ and $\boldsymbol{Y_2}$, respectively. After the posterior for each vector has been calculated then we can calculate the difference, this is possible because the posterior will for $\mu_i \; \forall i \in (1,2)$ will be a t-distribution.
(Reference: Bayesian Data Analysis Third edition, chapter: 3 , page: 66)

Because we know that the posterior distribution for μ_1 and μ_2 is t-distributed and the t-distribution is from the scale-location family then μ_d will also be t-distributed.

Likelihood function for y_1:
$L(y_1|\mu_1, \sigma_1^2) = \prod_{i=1}^{n_1} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2\sigma_1^2}\left(y_i - \mu_1\right)^2\right)$

Likelihood function for y_2:
$L(y_2|\mu_2, \sigma_2^2) = \prod_{i=1}^{n_2} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2\sigma_2^2}\left(y_i - \mu_2\right)^2\right)$

Noninformative prior distribution for y_1 and σ_1^2:
$P(\mu_1, \sigma_1^2) \propto \frac{1}{\sigma_1^2}$

Noninformative prior distribution for y_2 and σ_2^2:
$P(\mu_2, \sigma_2^2) \propto \frac{1}{\sigma_2^2}$

Joint posterior distribution for μ_1 and σ_1^2 given $\boldsymbol{Y_1}$:
$p\left(\mu_1, \sigma_1^2 \mid y_1\right) \propto \sigma^{-n_1-2} \exp\left(-\frac{1}{2\sigma_1^2} \sum_{i=1}^{n_1}\left(y_i - \mu_1\right)^2\right)$

Marginal posterior distribution for μ_1:
$p(\mu_1 \mid y_1) = t_{n_1-1}\left(\mu_1 \mid \bar{y}_1, s_1^2/n_1\right) = t_{9-1}\left(\mu_1 \mid \bar{y}_1, s_1^2/n_1\right)$

The same process is being done for calculating the μ_2 and we get the following marginal posterior distribution for μ_2 by integrate with respect to σ_2^2:
$p(\mu_2 \mid y_2) = t_{n_2-1}\left(\mu_2 \mid \bar{y}_2, s_2^2/n_1\right) = t_{n_{13}-1}\left(\mu_2 \mid \bar{y}_2, s_2^2/n_1\right)$

Because both μ_1 and μ_2 comes from the t-distribution and the t-distribution is from the scale-location family then we can take the difference between the parameters which still is a t-distribution.

Now we are creating a function that computes the point estimate for μ_d:

```r
mu_diff_point_est <- function(data1 = windshieldy1,
                              data2 = windshieldy2) {
  # the point estimate for mu_d where mu_d = mu_1 - mu_2
  point_est <- mean(windshieldy1) - mean(windshieldy2)

  return(point_est)
}
mu_diff_point_est()
```

```
## [1] -1.209855
```

and the point estimate for μ_d corresponds to -1.209855. Now we are creating a function that computes the 95% posterior interval for μ_d:

```r
mu_diff_posterior_interval <- function(data1 = windshieldy1,
                                       data2 = windshieldy2,
                                       prob = 0.95) {
  set.seed(1212)
  n1 <- length(data1)
  n2 <- length(data2)

  sd1 <- sd(data1) / sqrt(n1)
  sd2 <- sd(data2) / sqrt(n2)

  mean1 <- mean(data1)
  mean2 <- mean(data2)

  sample1 <- rtnew(100000, df = n1 - 1, mean = mean1, scale = sd1)
  sample2 <- rtnew(100000, df = n2 - 1, mean = mean2, scale = sd2)

  mu_d <- sample1 - sample2

  output <- round(quantile(mu_d,
                    probs = c( (1-prob)/2,
                               1-(1-prob)/2)), digits = 3)
  return(output)

}
mu_diff_posterior_interval()
```

```
##   2.5%  97.5%
## -2.448  0.025
```

and the 95 % posterior interval is $[-2.448, 0.025]$ such that
$P(0.321 < \mu_d < 0.925) \approx 0.95$. We are now plotting the histogram for μ_d according
to below:

```r
data1 <- windshieldy1
data2 <- windshieldy2

n1 <- length(data1)
n2 <- length(data2)
```
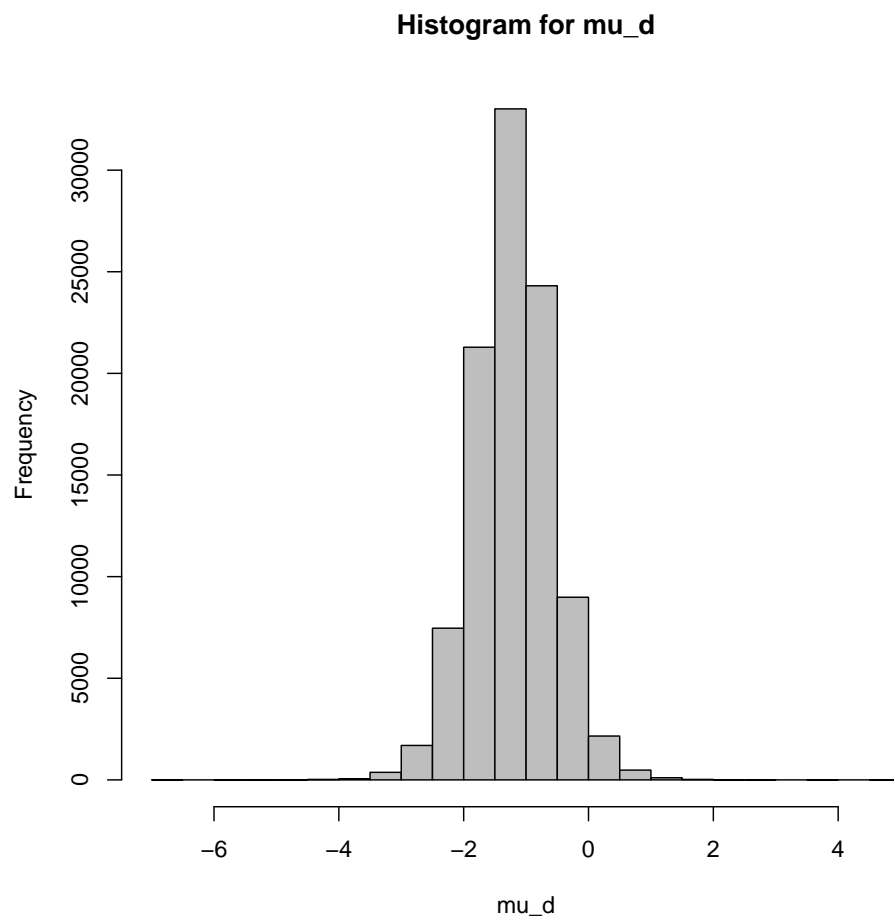
12

```
sd1 <- sd(data1) / sqrt(n1)
sd2 <- sd(data2) / sqrt(n2)

mean1 <- mean(data1)
mean2 <- mean(data2)

sample1 <- rtnew(100000, df = n1 - 1, mean = mean1, scale = sd1)
sample2 <- rtnew(100000, df = n2 - 1, mean = mean2, scale = sd2)
mu_d <- sample1 - sample2
hist(mu_d,
     main = "Histogram for mu_d",
     col = "gray")
```

Histogram for mu_d

Q3b, Probability of equal means

The probability that $\mu_1 = \mu_2$ corresponds to zero. Reason being that the posterior distributions of μ_1 and μ_2 are continuous hence the posterior distribution of the difference μ_d is also continuous and for a continuous random variable to be equal to a specific value always corresponds to zero. So $P(\mu_d = 0) = 0$ for the whole range, where $\mu_d = \mu_1 - \mu_2$.