CS-E5710 Bayesian Data Analysis

Assignment 3

Anomymous

1. Inference for normal mean and deviation

Following likelihood, prior, joint posterior, and marginal posterios distribution of μ are derived in the course text book[1].

- Model likelihood:

$$p(y \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)}$$

- Model prior (standard uninformative):

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$$

- Resulting joint posterior distribution:

$$p(\mu, \sigma^2 \mid y) \propto \sigma^{-n-2} exp(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y} - \mu)^2]), \text{ where}$$
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

- Marginal posterior for μ:

$$p(\mu|y) = \int_0^{\inf} p(\mu, \sigma^2|y)d\sigma^2$$
$$p(\mu|y) \propto \left[1 + \frac{n(\mu-\bar{y})^2}{(n-1)s^2}\right]^{-n/2}$$
This is the $t_{n-1}(\bar{y}, s^2/n)$ density.

a) What can we say about the unknown μ?

We are interested in Bayesian point estimate for the mean of μ, 95% posterior interval, and density plot. We approach the problem by sampling from marginal distribution:

```
mu_point_est <- function(data){
  mean(data)
  n = length(data)
  mu = mean(data)
  std = sd(data)
  n_samples = 1000000
  samples = rtnew(n_samples,n-1,mean=mu,scale=std/sqrt(n))
  point_est = mean(samples)
}

mu_interval<-function(data , prob ){
  n = length(data)
  std = sd(data)
  mu = mean(data)
  n_samples = 1000000
  lower = (1 - prob) / 2
```

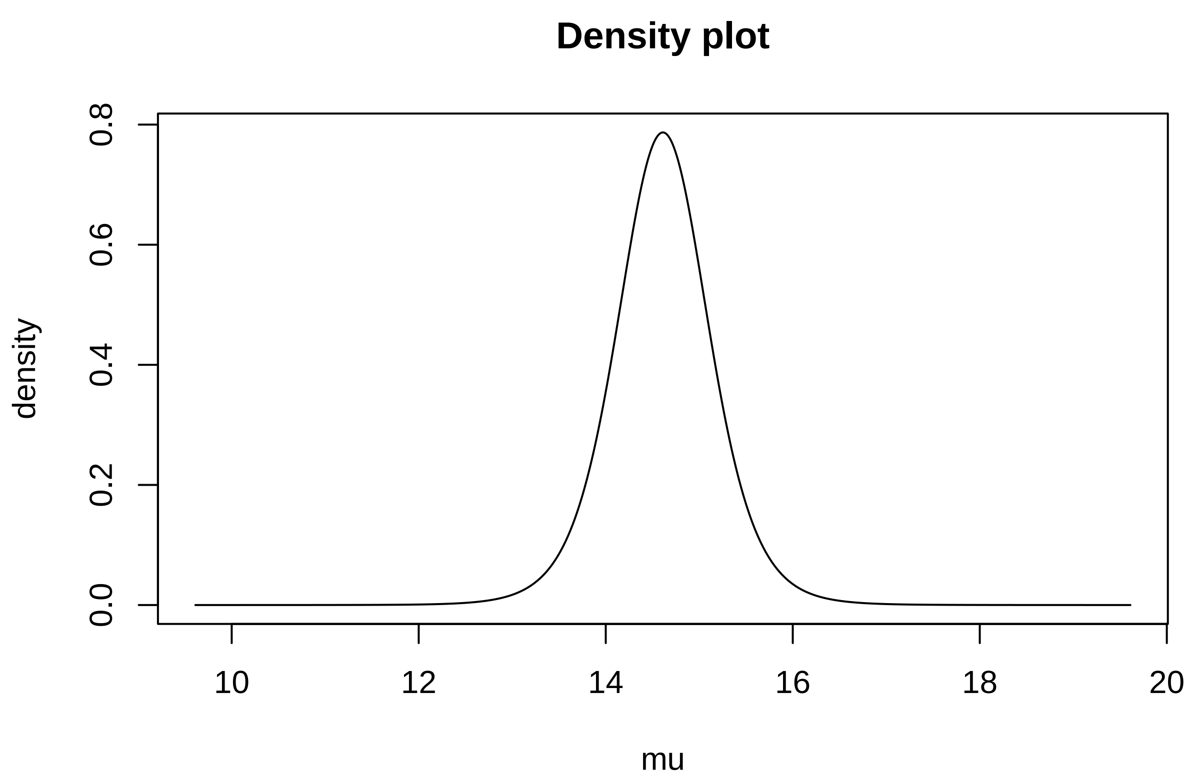
```r
  upper = 1- lower
  samples = rtnew(n_samples,n-1,mean=mu,scale=std/sqrt(n))
  interval <- quantile(samples,c(lower,upper))
}

plot_density <- function(data){
  n = length(data)
  mu = mean(data)
  std = sd(data)
  x = seq(mu-5, mu+5, 1/1000)
  pdf = dtnew(x, n-1, mean = mu, scale = std/sqrt(n))
  plot(x, pdf, type="l", xlab ="mu" ,ylab = "density", main = "Density plot")
}

mu_point = mu_point_est(windshieldy1)
sprintf("Bayesian point estimate for mu: %.2f",mu_point)
```

```
## [1] "Bayesian point estimate for mu: 14.61"
```

```r
mu_int = mu_interval(data = windshieldy1, prob=0.95)
sprintf("The posterior 95 percent interval for mu %.2f : %.2f",mu_int[1],mu_int[2])
```

```
## [1] "The posterior 95 percent interval for mu 13.48 : 15.74"
```

```r
plot_density(windshieldy1)
```



b) What can we say about the hardness of the next windshield coming from the production line?

Now we sample $\bar{y}$ from posterior predictive distribution. Using the same uninformative prior distribution as in part a), the posterior predictive distribution is the $t_{n-1}(\bar{y}, s\sqrt{1+\frac{1}{n}})$ density, as given in the course text book [1].

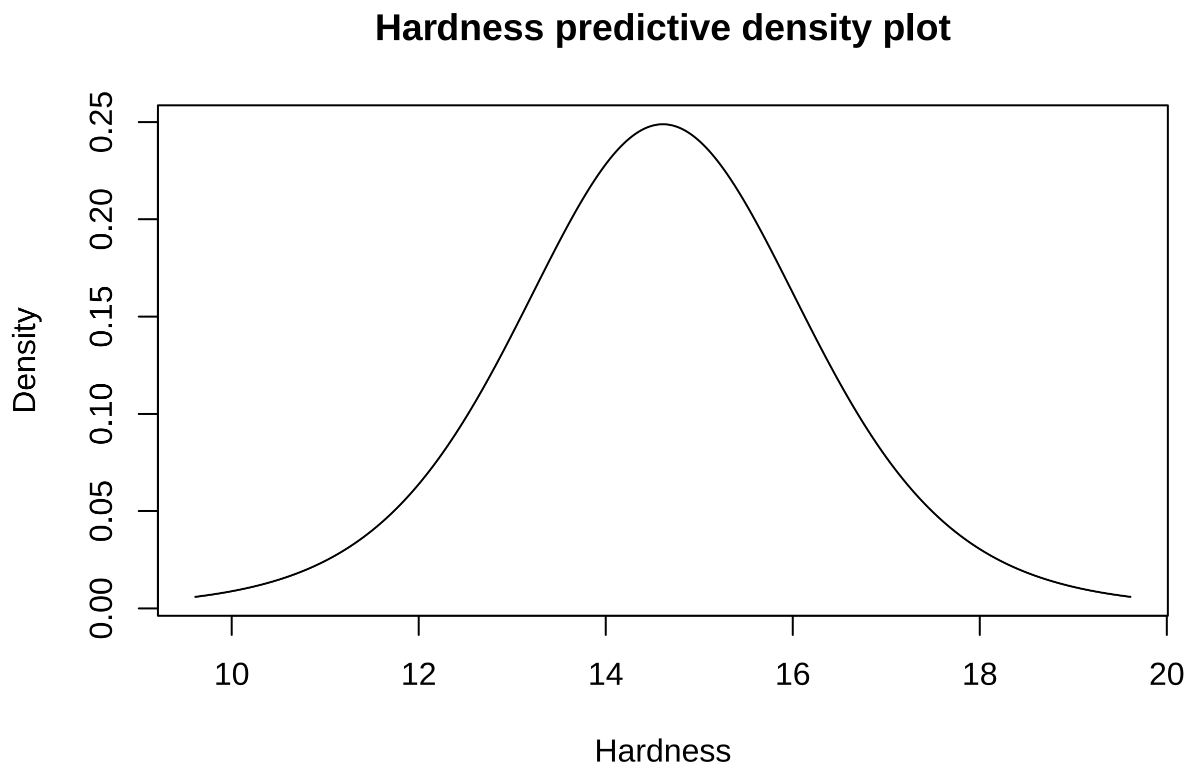
```
mu_pred_point_est<-function(data){
  mean(data)
  n = length(data)
  mu = mean(data)
  std = sd(data)
  n_samples = 1000000
  samples = rtnew(n_samples,n-1,mean=mu,scale=(std*sqrt((1 + 1/n))))
  point_est = mean(samples)
}

pred_mu_interval<-function(data, prob){
  n = length(data)
  mu = mean(data)
  std = sd(data)
  lower = (1 - prob) / 2.0
  upper = 1- lower
  borders = qtnew(c(lower , upper), n-1, scale = (std*sqrt((1 + 1/n))))
  borders + mu
}

plot_density_b <- function(data){
  n = length(data)
  mu = mean(data)
  s = sd(data)
  seq = seq(mu-5, mu+5, 1/1000)
  pdf = dtnew(seq, n-1, mean = mu, scale=s*sqrt((1+ 1 /n)))
  plot(seq, pdf, type="l", xlab ="Hardness" ,ylab = "Density",main="Hardness predictive density plot")
}

mu_pred_point_estimate = mu_pred_point_est(data=windshieldy1)
mu_intervall_est= pred_mu_interval(data = windshieldy1, prob=0.95)
sprintf("The point estimate for hardnessis %.2f", mu_pred_point_estimate)
```

```
## [1] "The point estimate for hardnessis 14.61"
```

```
sprintf("The 95 percent predictive interval for hardness %.2f : %.2f",mu_intervall_est[1], mu_intervall_
```

```
## [1] "The 95 percent predictive interval for hardness 11.03 : 18.19"
```

```
plot_density_b(windshieldy1)
```



2. Inference for the difference between portions

From the course textbook [1] we get the following formulas for binomial model:

- Likelihood $= p(y|\theta) \propto \theta^a(1-\theta)^b$
- Prior $= p(\theta) = Beta(y|\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$
- Posterior $= p(\pi|y) \propto \theta^y(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1}$
  $= \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$
  $= Beta(\theta|\alpha+y, \beta+n-y)$

Using the formulas above and values for both groups (0=Control, 1=Treatment) we get following:

(1) Likelihood: $p(y_0|p_0) \sim Binomial(n_0, p_0) = Binomial(39, 674)$ and $p(y_1|p_1) \sim Binomial(n_1, p_1) = Binomial(22, 680)$

(2) Uninformative prior (none tested) for both groups: $p(p_0), p(p_1) \sim Beta(1|1)$

(3) Posterior: $p(p_0|y_0) \sim Beta(1+y_0, 1+n_0-y_0) = Beta(40, 636)$ and $p(p_0|y_0) \sim Beta(1+y_0, 1+n_0-y_0) = Beta(23, 659)$

a) Summarize the posterior for the odds ratio $\frac{p_1}{1-p_1}/\frac{p_0}{1-p_0}$.

We sample from both posterior distributions, and calculate odds-ratios. For the posterior odds-ratio, we calculate Bayesian point estimate for odds-ratio mean, 95 percent intervall, and visualize the odds-ratios.

```
#uniformative prior parameters
prior_a = 1 # Success
prior_b = 1 # Total -success
# Control
```

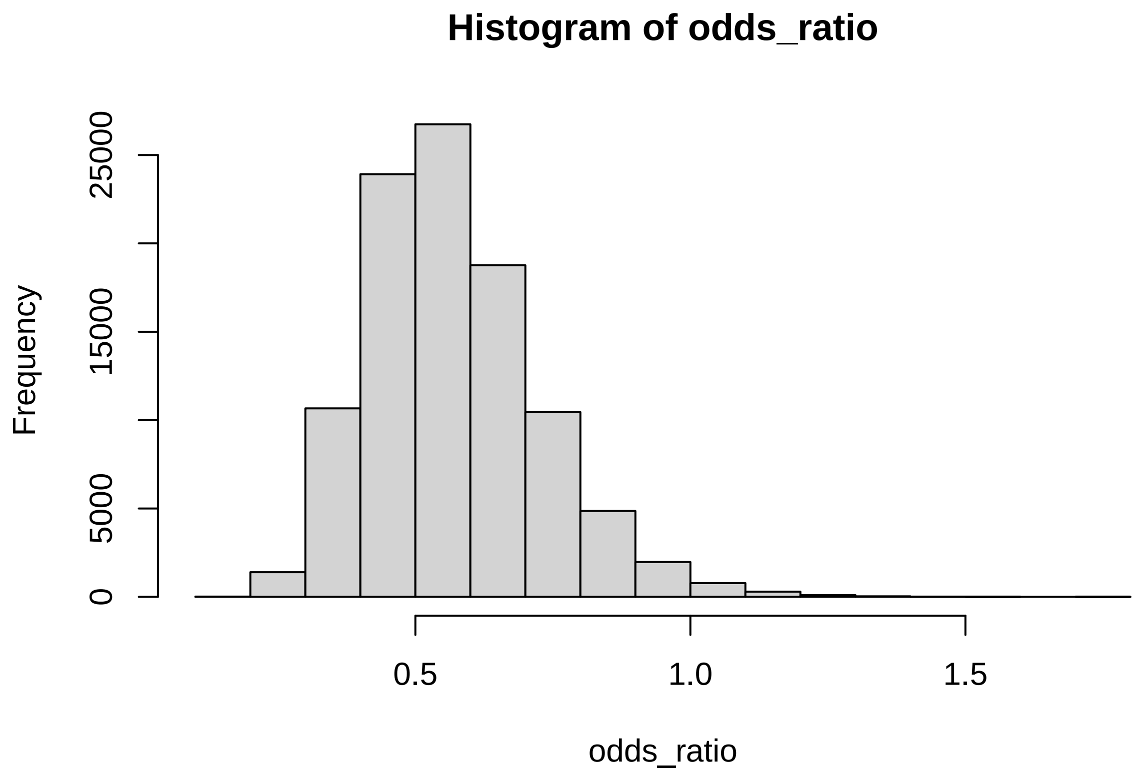
```r
y_1 = 39
n_1 = 674
# Treatment
y_2 = 22
n_2 = 680
n = 100000
p0 = rbeta(n,(prior_a+y_1),(prior_b+n_1-y_1))
p1 = rbeta(n,(prior_a+y_2),(prior_b+n_2-y_2))

posterior_odds_ratio_point_est <- function(p0,p1){
  mean((p1/(1-p1))/(p0/(1-p0)))
}

posterior_odds_ratio_interval <- function(p0,p1,prob){
  n = length(p0)
  samples = (p1/(1-p1)) / (p0/(1-p0))
  lower <- (1 - prob) / 2
  upper <- 1 - lower
  limits = c(lower,upper)
  quantile(samples, limits)
}

p_est = posterior_odds_ratio_point_est(p0,p1)
sprintf("The point estimate for odds ratio is %.2f", p_est)
```

```
## [1] "The point estimate for odds ratio is 0.57"
```

```r
p_odds_int = posterior_odds_ratio_interval(p0,p1,prob=0.95)
sprintf("The borders of 95 percent %% of the estimated intervalls are %.2f : %.2f", p_odds_int[1],p_odds
```

```
## [1] "The borders of 95 percent % of the estimated intervalls are 0.32 : 0.93"
```

```r
odds_ratio = ((p1/(1-p1)) / (p0/(1-p0)))
hist(odds_ratio)
```




```
prob_better = mean(p1-p0 < 0)
sprintf("The probability that Treatment yields lower mortality is %.4f", prob_better)
```

```
## [1] "The probability that Treatment yields lower mortality is 0.9878"
```

Assuming non-informative prior, the probability that beta-blockers given to Treatment group lowers the mortality when compared to Control group is 98,8 %.

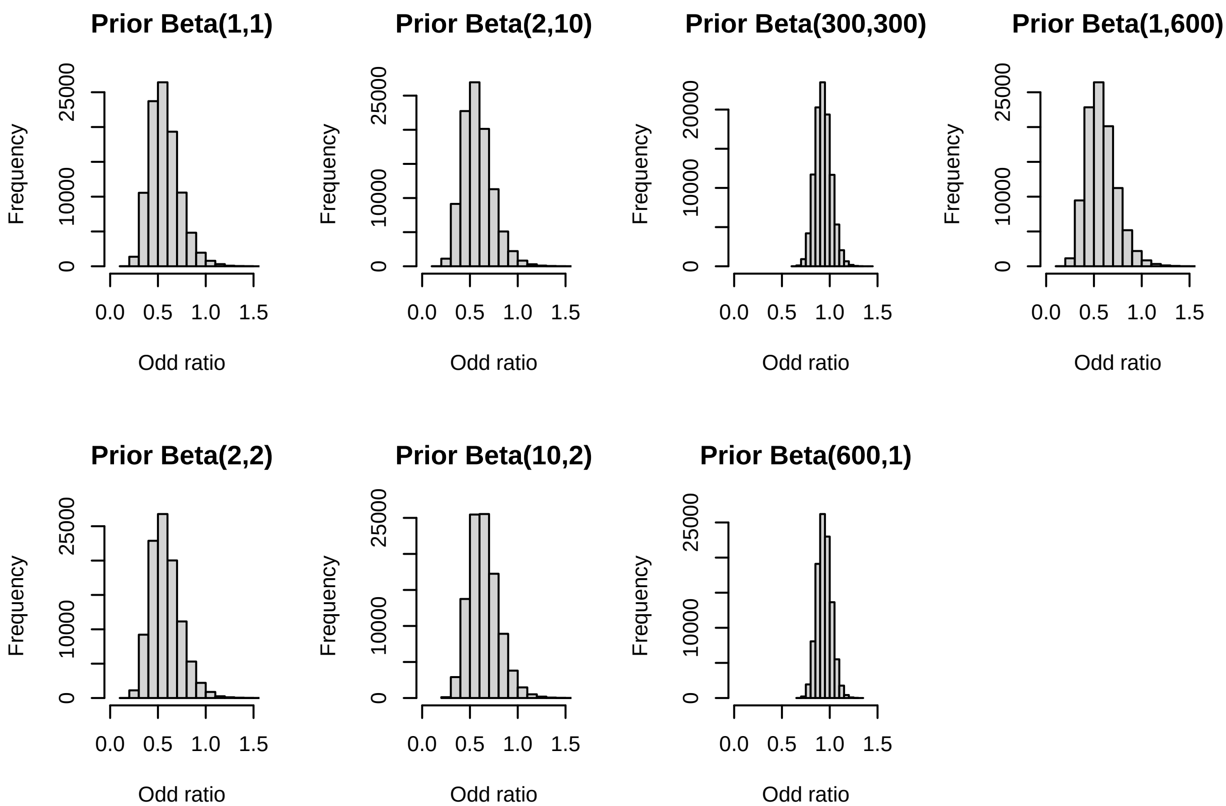
b) Discuss the sensitivity of your interference to your choice of prior density with couple of sentences

In part a) we had uninformative prior $Beta(1, 1)$. Here we examine the effect of prior, choosing weakly informative prior $Beta(2, 2)$, informative priors $Beta(2, 10)$, $Beta(10, 2)$, $Beta(300, 300)$, $Beta(600, 1)$, $Beta(1, 600)$.

```
df <- data.frame(matrix(ncol = 6, nrow = 0))
x <- c("prior_alpha", "prior_beta", "point_estimate","low_interval","high_interval","prob_better")
colnames(df) <- x
df[nrow(df) + 1,] <- c(prior_a, prior_b, p_est, p_odds_int[1],p_odds_int[2], prob_better)

a = c(1,2,2,10,300,600,1)
b = c(1,2,10,2,300,1,600)

par(mfcol=c(2,4))

for(i in 1:7){
  prior_a = a[i]
  prior_b = b[i]
```

```
  p0 = rbeta(n,(prior_a+y_1),(prior_b+n_1-y_1))
  p1 = rbeta(n,(prior_a+y_2),(prior_b+n_2-y_2))
  p_est = posterior_odds_ratio_point_est(p0,p1)
  p_odds_int = posterior_odds_ratio_interval(p0,p1,prob=0.95)
  odds_ratio = ((p1/(1-p1)) / (p0/(1-p0)))
  hist(odds_ratio,xlim=c(0, 1.5),xlab="Odd ratio",main=paste("Prior Beta(",prior_a,",",prior_b,")", sep
  prob_better = mean(p1-p0 < 0)

  df[nrow(df) + 1,] <- c(prior_a, prior_b, p_est, p_odds_int[1],p_odds_int[2], prob_better)
}

print(df)
```

```
##   prior_alpha prior_beta point_estimate low_interval high_interval prob_better
## 1           1          1      0.5693625    0.3209788     0.9255834     0.98784
## 2           1          1      0.5706196    0.3216391     0.9246827     0.98775
## 3           2          2      0.5795040    0.3292545     0.9373193     0.98693
## 4           2         10      0.5797488    0.3285369     0.9349371     0.98714
## 5          10          2      0.6446470    0.3920837     0.9910820     0.97732
## 6         300        300      0.9304654    0.7758271     1.1062331     0.80080
## 7         600          1      0.9423660    0.8039357     1.0968558     0.78547
## 8           1        600      0.5793119    0.3282647     0.9362945     0.98643
```



Based on prior sensitivity analysis, we notice that using informative priors $Beta(300, 300)$ (high mortality) and $Beta(600, 1)$ (even higher mortality) yield differing results, in comparison with other priors. Odds-ratio does not seem to be very sensitive to small changes in prior parameters, but is affected by larger values indicating higher baseline mortality. This is due that the data overweigths the small prior parameters.

3. Interference for the difference between normal means

As in exercise 1., likelihood, prior, joint posterior, and marginal posteriors distribution of μ are derived in the course text book[1].

- Model likelihood:

$$p(y \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)}$$

- Model prior (standard uninformative):

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$$

- Resulting joint posterior distribution:

$$p(\mu, \sigma^2 \mid y) \propto \sigma^{-n-2} exp(-\tfrac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]), \text{ where}$$
$$s^2 = \tfrac{1}{n-1}\textstyle\sum_{i=1}^{n}(y_i - \bar{y})^2$$

- Marginal posterior for μ:

$$p(\mu|y) = \int_0^{\inf} p(\mu, \sigma^2|y) d\sigma^2$$
$$p(\mu|y) \propto \left[1 + \frac{n(\mu-\bar{y})^2}{(n-1)s^2}\right]^{-n/2}$$
This is the $t_{n-1}(\bar{y}, s^2/n)$ density.

a) what can we say about the difference $\mu_d = \mu_1 - \mu_2$?

We assume that production lines are independent of each other, and we can utilize exchangeability. Thus, we compare the means by sampling from both posteriors and compare the observation differences:

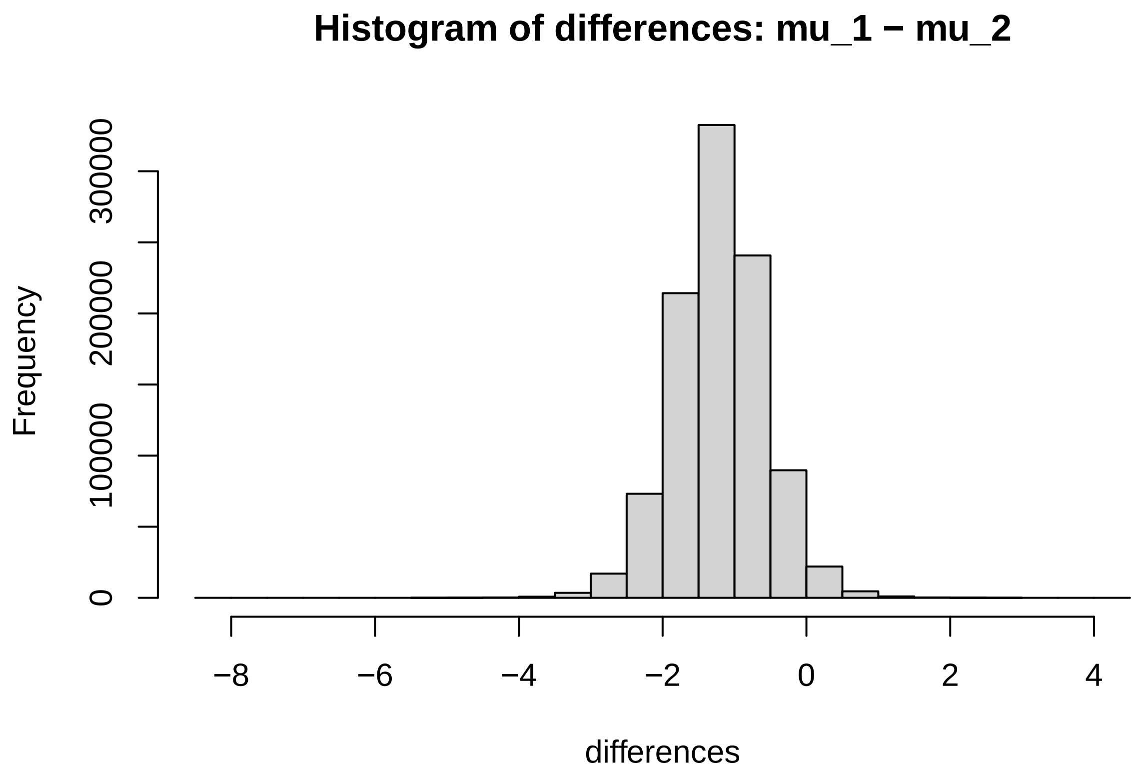
```
windshield_sample <- function(data){
  n = length(data)
  mu = mean(data)
  std = sd(data)
  n_samples = 1000000
samples = rtnew(n_samples,n-1,mean=mu,scale=std/sqrt(n))
}

mu_1 = windshield_sample(data=windshieldy1)
mu_2 = windshield_sample(data=windshieldy2)

differences = mu_1 - mu_2
mean_diff = mean(differences)


posterior_diff_interval <- function(differences,prob){
  n = length(p0)
  lower <- (1 - prob) / 2
  upper <- 1 - lower
  limits = c(lower,upper)
  quantile(differences, limits)
}

mu_interval = posterior_diff_interval(differences,p=0.95)

sprintf("The Bayesian point estimate for difference mu_1 - mu_2 is %.2f", mean_diff)
```

```
## [1] "The Bayesian point estimate for difference mu_1 - mu_2 is -1.21"
```

```
sprintf("The estimated 95 percent posterior interval for the difference is %.2f : %.2f", mu_interval[1]
```

```
## [1] "The estimated 95 percent posterior interval for the difference is -2.45 : 0.04"
```

```
hist(differences,main="Histogram of differences: mu_1 - mu_2")
```




```
prob_thicker = mean(mu_1 - mu_2 < 0)
sprintf("The probability that production line2 produces thicker glass is %f", prob_thicker)
```

```
## [1] "The probability that production line2 produces thicker glass is 0.972110"
```

Assuming uninformative prior, the probability that production line 2 produces thicker glass when compared to production line 1 is 97,2 %.

b) What is the probability that the means are exactly the same ($\mu_1 - \mu_2$)?

There is very low probability that $\mu_1 > \mu_2$, but the probalibility that $\mu_1 = \mu_2$ is 0. This is because posterior distributions of μ_1 and μ_2, and their difference $\mu_d = \mu_1 - \mu_2$ are continuous, thus the probability that difference having any particular value is 0.

References:

1. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data Analysis (3rd ed.). Chapman & Hall/CRC.