

Classical and Bayesian Statistics

Problems 13

Note that you have to run the following code *once* first:

```
library(rstan)
library(bayestestR)
library(latex2exp)
rstan_options(auto_write = TRUE)

plot_posterior <- function(params, Rope = c(-100, -99)){
  dens <- density(params$mu)
  max_dens <- max(dens$y)
  hdi_l <- as.numeric(hdi(params$mu)[2])
  hdi_r <- as.numeric(hdi(params$mu)[3])

  plot(dens, main="Posterior", col="darkseagreen", TeX("$\\mu$"), lwd=2)

  lines(c(hdi_l, hdi_r), c(.12*max_dens, .12*max_dens), col="orange", lwd=2)
  text(hdi_l, .15*max_dens, round(hdi_l, 3), cex=.8)
  text(hdi_r, .15*max_dens, round(hdi_r, 3), cex=.8)
  text((hdi_l+hdi_r)/2, .15*max_dens, "95% HDI", cex=.8, col="orange")

  lines(c(Rope[1], Rope[2]), c(.08*max_dens, .08*max_dens), col="blue", lwd=2)
  text(Rope[1], .05*max_dens, Rope[1], cex=.8)
  text(Rope[2], .05*max_dens, Rope[2], cex=.8)
  text((Rope[1]+Rope[2])/2, .05*max_dens, "ROPE", cex=.8, col="blue")
}
```

Problem 13.1

A wine merchant claims that the wine bottles he fills contain 70 centilitres. However, a sceptical consumer suspects that the wine merchant does not fill the bottles with enough wine and wants to “check” this claim. Therefore, he buys 12 bottles of wine and measures their contents. He finds:

71, 69, 67, 68, 73, 72, 71, 71, 68, 72, 69, 72 (in centiliters)

First, assume that the standard deviation of the bottling is known in advance. It is $\sigma = 1.5$ centilitres.

- a) Check whether a normal distribution assumption is justified. Justify your answer!
- b) We choose a uniform prior distribution and express a high degree of uncertainty because we do not trust the wine merchant. How would you choose the boundaries of the uniform distribution?
- c) Determine the 95 %-HDI and the mode of the data.
- d) We know that the wine bottles are filled with very high accuracy. Choose an appropriate ROPE.
- e) How do you interpret the result? Decide whether the wine merchant has under-filled the bottles.
- f) The standard deviation was assumed to be given. What is the result of d) if we estimate the standard deviation from the data?

Problem 13.2

Two depth gauges measure the following values for the depth of a rock layer at nine different locations:

Gauge A	120	265	157	187	219	288	156	205	163
Gauge B	127	281	160	185	220	298	167	203	171

It is suspected that device *B* systematically measures larger values. Do the measurements confirm this suspicion or is a random fluctuation plausible as an explanation?

- a) Are the samples paired or independent?
- b) Is the assumption of a normal distribution for the difference justified?
- c) Choose a uniform prior distribution for μ of the difference and choose the prior for the standard deviation using the standard deviation of the data.
- d) Is it true that gauge *B* differs significantly from gauge *A*? To do this, choose an appropriate ROPE.
- e) How do you interpret the result?

Male x_i	120	107	110	116	114	111	113	117	114	112
Female y_j	110	111	107	108	110	105	107	106	111	111

Problem 13.3

The following table shows the jaw lengths of 10 male and 10 female golden jackals: Some key figures: $\bar{x}_n = 113.4$, $\bar{y}_n = 108.6$, $\hat{\sigma}_x^2 = 13.82$, $\hat{\sigma}_y^2 = 5.16$. The question now arises as to whether there is a difference in jaw length between males and females.

```
data <- read.table("jackals.txt", sep=" ", header = T)
head(data)

data_m <- data[, "M"]
data_f <- data[, "W"]
```

- Are the samples paired or unpaired? Justify your answer.
- Is the normal distribution assumption justified for the data?
- Choose an appropriate uniform prior distribution for μ . Choose the estimated standard deviation.
- Is there a statistically significant difference between the jaw lengths of male and female jackals? How do you interpret the result?

Note: Proceed in the same way as in the R-file [student_heights.R](#)

Problem 13.4

A U.S. magazine, Consumer Reports, conducted a study of the calorie and sodium content of various hotdog brands. There were three different types of hot dogs: beef, „meat“ (beef, pork, poultry mix), and poultry.

The results below list the calorie content of various brands of beef and poultry hotdogs.

Beef hotdog

186, 181, 176, 149, 184, 190, 158, 139, 175, 148, 152, 111, 141, 153, 190, 157, 131, 149, 135, 132

Poultry hotdog:

129, 132, 102, 106, 94, 102, 87, 99, 170, 113, 135, 142, 86, 143, 152, 146, 144

Do the two hotdog types have different calorie content?

- Are the samples paired or unpaired? Justify your answer.
- Is the normal distribution assumption justified for the data?
- Choose an appropriate uniform prior distribution.
- Is there a statistically significant difference between the calorie contents? Use the estimated standard deviation. How do you interpret the result?

Problem 13.5

In 2013, as part of an international collaboration led by EAWAG in Dübendorf, concentrations of illegal substances in the wastewater of 42 European cities were examined over the course of a week (Ort C. et al, *Spatial differences and temporal changes in illicit drug use in Europe quantified by wastewater analysis*, Addiction 2014 Aug).

In this study, the median concentrations of ecstasy (MDMA) in wastewater were measured along with other substances on seven consecutive days (March 6-12). Based on this study, a widely read Swiss free newspaper claimed that many more drugs are consumed in Zurich than elsewhere.

The following table shows the quantities of MDMA excreted on the days of the study for the cities of Zurich and Basel. The data are given in mg per 1000 inhabitants per day. Assume that the daily differences D_i between the quantities of MDMA excreted

Weekdays	We	Th	Fr	Sa	Su	Mo	Tu
Zurich	16.3	12.7	14.0	53.3	117	62.6	27.6
Basel	10.4	8.91	11.7	29.9	46.3	25.0	29.4

per thousand inhabitants in the wastewater of Zurich and Basel are independently of each other normally distributed with an expected value μ_D and a standard deviation σ_D .

- Estimate the mean and the standard deviation of the differences, i.e., $\hat{\mu}_D$ and $\hat{\sigma}_D$, from the data (D stands for difference).
- Is the sample paired or unpaired? Justify your answer.
- Is the assumption of normality justified for the data?
- Choose an appropriate uniform prior distribution. Use the estimated standard deviation and choose an appropriate ROPE.

- e) Is there a statistically significant difference between the MDMA use in the two cities? How do you interpret the result?

Problem 13.6

In the R-file `husband_wife.R` we have already looked at the age difference between married couples. Now we want to examine the height difference. According to Wikipedia, in England men are taller than women by 13 cm (in general and not just for married couples). We want to investigate whether this also applies to married couples, namely that height difference of husband and wife is also 13 cm.

The data set `husband_wife.csv` contains the values of height and age for 170 British married couples. The height is given in cm and the age in years.

Note:

- We can argue for paired or unpaired samples. We consider paired samples. What is the interpretation of a (possible) height difference of 13 cm?
- Is the assumption of a normal distribution justified for the data? Choose an appropriate uniform prior distribution. Use the estimated standard deviation for this.
- Is there a statistically relevant difference compared to the difference in the general population? How do you interpret the result?

Problem 13.7

The body temperature of 10 patients is measured at the time of administration of a drug (T_1) and 2 hours later (T_2). It should be examined whether this drug has an antipyretic effect.

Patient-Nr.	1	2	3	4	5	6	7	8	9	10
Temp. 1 in °C	39.1	39.3	38.9	40.6	39.5	38.4	38.6	39.0	38.6	39.2
Temp. 2 in °C	38.1	38.3	38.8	37.8	38.2	37.3	37.6	37.8	37.4	38.1

- Is it a paired or unpaired test? Justify your answer.
- Is the normal distribution assumption justified for the data?

- c) Choose a reasonable uniform prior distribution for μ . Use the estimated standard deviation.
- d) Is there a statistically significant difference between the body temperatures of T_1 and T_2 ? Choose a reasonable ROPE. How do you interpret the result?

Classical and Bayesian Statistic

Sample solution for Problems 13

Solution 13.1

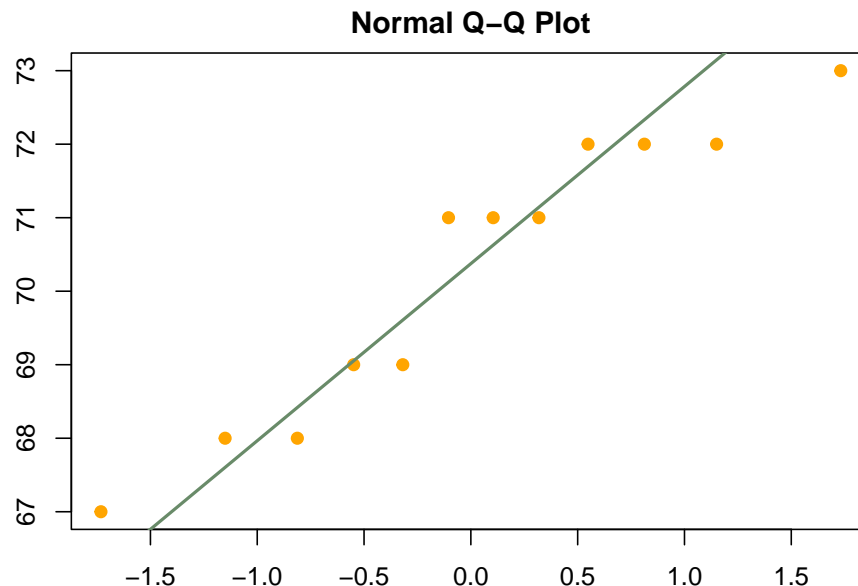
We use the following notation:

X_i = content (in centiliters) of the i th wine bottle, $i = 1, \dots, n = 12$.

- a) We now want to test the following assumption: X_1, \dots, X_{12} i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$, where $\sigma^2 = 1.5^2$ is known. To do this, we create a QQ plot:

```
data <- c(71, 69, 67, 68, 73, 72, 71, 71, 68, 72, 69, 72)

qqnorm(data, pch=19, col="orange")
qqline(data, col="darkseagreen4", lwd=2)
```



The QQ plot does not rule out a normal distribution. Since the data set is very small, a conclusive assessment is difficult.

- b) For the prior distribution, we choose a uniform distribution. We specify the right end of the uniform distribution as 80, since often a little more fits into the bottle.

```
modelString = "
data {
  int n; # Number of data points
  real y[n]; # Data points are real numbers
```

```

}

parameters {
  real<lower=0,upper=80> mu; # Uniform prior distribution for mu
  real<lower=1.5,upper=1.6> sigma; # Uniform prior for sigma (constant)
}

model {
  for (i in 1:n)
    y[i] ~ normal(mu,sigma); # Likelihood function for data
}
" # close quote for modelString

model <- stan_model(model_code=modelString)

n      = length(data)
y      = data

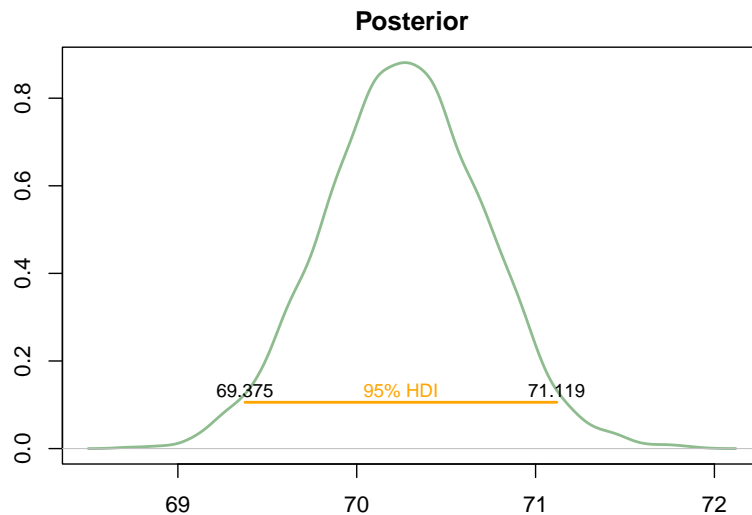
stanFit = sampling( object=model , data=list( y = y , n = n ))

params = rstan::extract(stanFit)

plot_posterior(params)

```

c) Posterior distribution:



The mode is about 70.3 which is quite interesting because we suspect that the

mean content of the bottles is *less than* 70 cl. A look at the mean of the data confirms our observation:

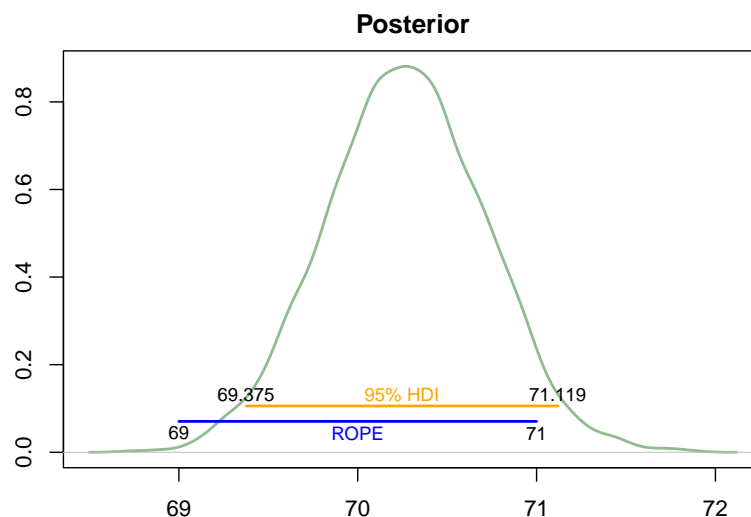
```
mean(data)
```

```
70.25
```

Because the prior distribution is quite wide, the posterior distribution reflects the data.

The HDI is from 69.4 to 71.1.

- d) We choose the ROPE from 69 to 71. That is, we consider all values in this range to be equivalent to the null value 70.
- e) The posterior distribution with the ROPE drawn in



- f) Since most of the HDI is contained in the ROPE, we tend to accept the null value. Strictly speaking, we would have to refuse the decision because a (small) part of the HDI lies outside the ROPE.

- g) `sd(data)`

```
1.96
```

```
modelString = "  
data {  
  int n;  
  real y[n];
```

```
}

parameters {
  real<lower=0,upper=80> mu;
  real<lower=1.96,upper=2> sigma; # constant
}

model {
  for (i in 1:n)
    y[i] ~ normal(mu,sigma);
}
" # close quote for modelString

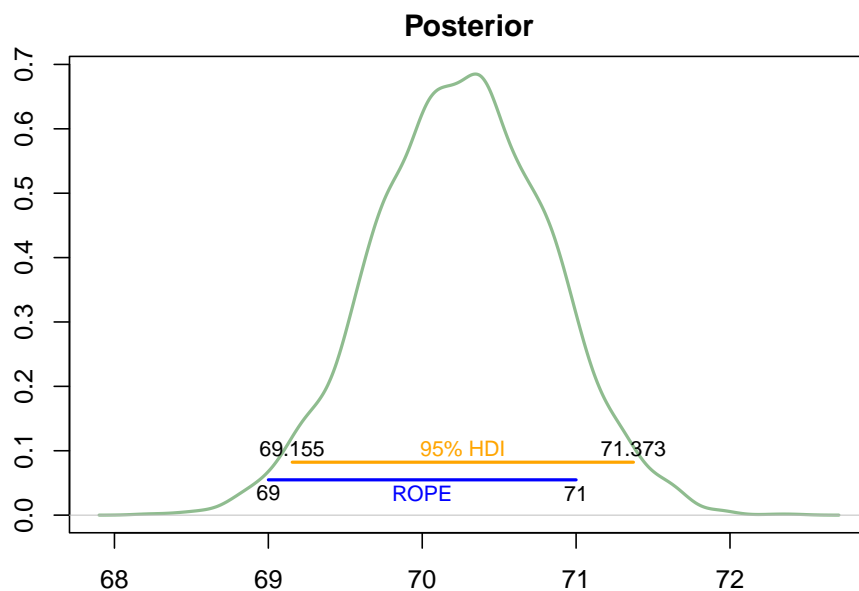
model <- stan_model(model_code=modelString)

n      = length(data)
y      = data

stanFit = sampling( object=model , data=list( y = y , n = n ))

params = rstan::extract(stanFit)

plot_posterior(params, Rope=c(69,71))
```



Now a larger part of the HDI lies outside the ROPE. In this case, we certainly

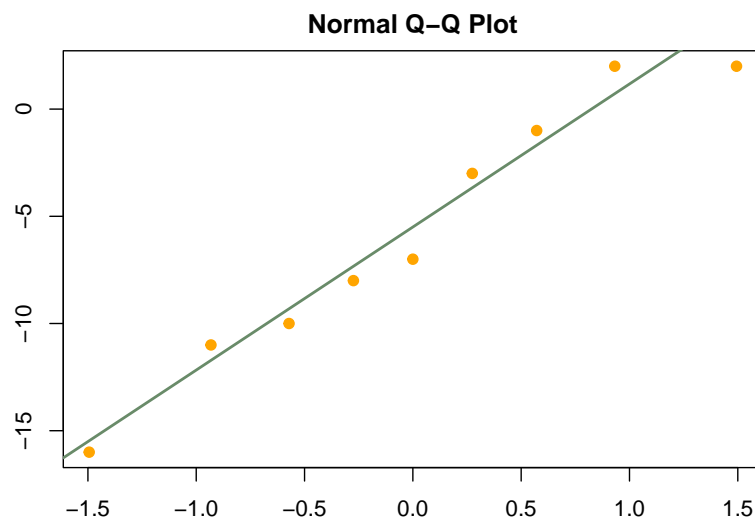
cannot reject the null value of 70.

Solution 13.2

- a) These are *paired* samples. The measurements are taken at the *same* location using both devices.
- b) To do this, we create a QQ plot:

```
df_A <- c(120, 265, 157, 187, 219, 288, 156, 205, 163)
df_B <- c(127, 281, 160, 185, 220, 298, 167, 203, 171)
data <- df_A - df_B

qqnorm(data, pch=19, col="orange")
qqline(data, col="darkseagreen4", lwd=2)
```



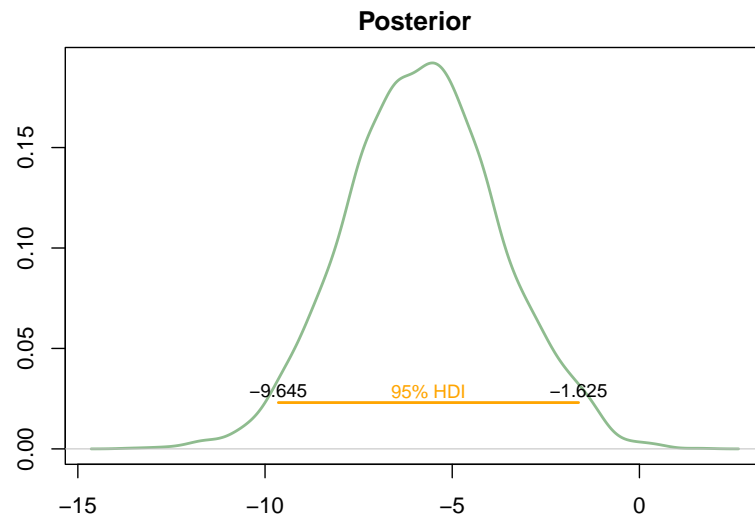
The QQ plot indicates a normal distribution. However, since the scope of the data set is very small, a conclusive assessment is difficult.

- c) We choose a uniform distribution between -20 and 20 for the differences.

The standard deviation is

```
sd(data)

6.200358
```



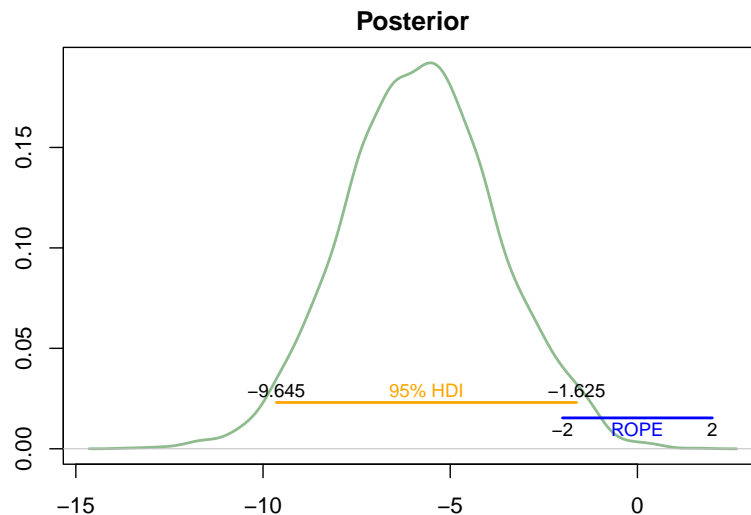
The mode of the posterior distribution is given as -5.7 and the 95 %-HDI between -9.6 and -1.6 .

- d) We choose the rope around zero because this does indicate no difference between the two gauge.

```
modelString = "  
data {  
  int n;  
  real y[n];  
}  
  
parameters {  
  real<lower=-20,upper=20> mu;  
  real<lower=6.2,upper=6.3> sigma;  
}  
  
model {  
  for (i in 1:n)  
    y[i] ~ normal(mu,sigma);  
}  
" # close quote for modelString  
  
model <- stan_model(model_code=modelString)  
  
n      = length(data)  
y      = data  
  
stanFit = sampling( object=model , data=list( y = y , n = n ))
```

```
params = rstan::extract(stanFit)

plot_posterior(params, Rope=c(-2,2))
```



- e) If the rope is chosen as $[-2, 2]$, then the rope is almost entirely outside the HDI, and we can assume that the two gauges provide statistically relevant different values. However, if we choose a larger rope, we would not be able to make a decision based on the data.

Solution 13.3

- a) These are unpaired samples, since no particular female belongs to the individual males. The sizes of the two samples do not have to be the same.

Therefore, we *cannot* use the difference of the two samples.

- b) We introduce the following designations:

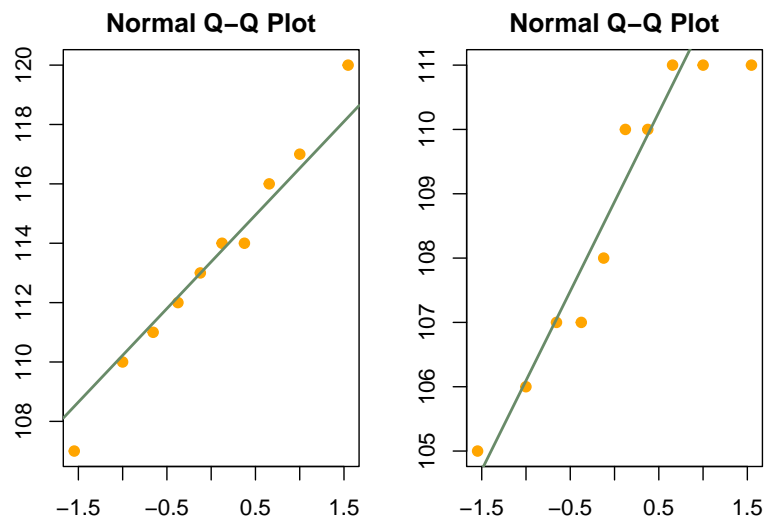
X_i : i th value of the length of the jaw of the male animals, $i = 1, \dots, n = 10$

Y_j : j th value of the length of the jaw of the female animals, $j = 1, \dots, m = 10$

```
par(mfrow=c(1,2)) # Divides graphical window in two (one row, two columns)

qqnorm(data_m, pch=19, col="orange")
qqline(data_m, col="darkseagreen4", lwd=2)

qqnorm(data_f, pch=19, col="orange")
qqline(data_f, col="darkseagreen4", lwd=2)
```



At first glance, the QQ plot of the males indicates a normal distribution, while that of the females does not. Since the scope of the data set is very small, a conclusive assessment is difficult.

- c) We choose for both a wide uniform prior distribution, namely from 50 to 200.

For the standard deviation

```
sd(data_m)
```

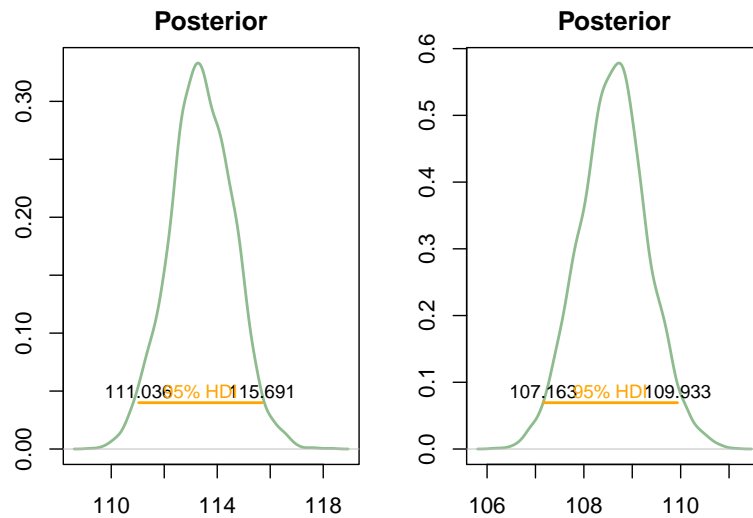
```
3.717825
```

```
sd(data_f)
```

```
2.270585
```

- d) The two HDIs do not overlap, and the HDI of the females is clearly separated from that of the males. Thus, we can assume that the jaw lengths are statistically

relevant between the sexes.



Solution 13.4

- a) We cannot clearly assign the measured values of one data set to the values of the other measured values. So this is an unpaired test. In addition, the series of measurements are of different lengths.

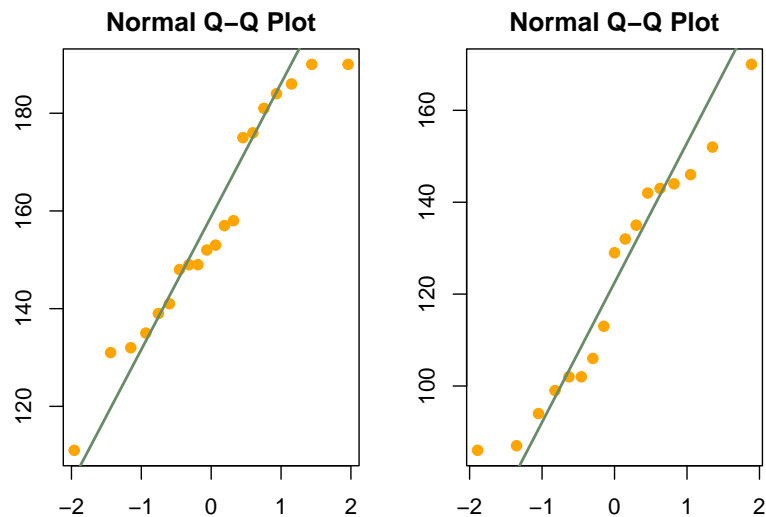
So, we *cannot* use the difference of the samples.

- b) Both QQ-Plots indicate a normal distribution. Since the data set is very small, a conclusive assessment is difficult.

```
par(mfrow=c(1,2)) # Divides graphical window in two (one row, two columns)

qqnorm(data_b, pch=19, col="orange")
qqline(data_b, col="darkseagreen4", lwd=2)

qqnorm(data_c, pch=19, col="orange")
qqline(data_c, col="darkseagreen4", lwd=2)
```



- c) We choose uniform distributions with boundaries from 50 to 250 as prior distributions.

For the standard deviation

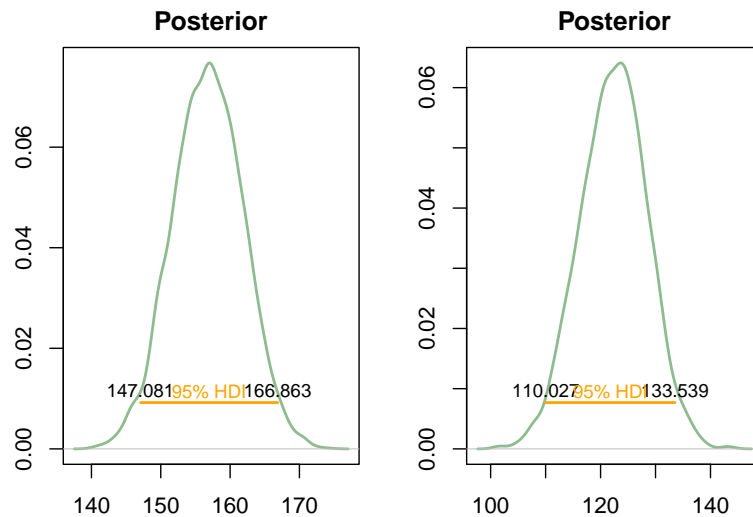
```
sd(data_b)
```

```
22.64201
```

```
sd(data_c)
```

```
25.48313
```

- d) The two 95 %-HDIs do not overlap. Thus, the beef hotdog has a statistically significant higher calorie content than the chicken hotdog.



Solution 13.5

- a) We estimate the mean and standard deviation with **Python** as follows:

```
mdma_zuerich <- c(16.3, 12.7, 14.0, 53.3, 117, 62.6, 27.6)
mdma_basel <- c(10.4, 8.91, 11.7, 29.9, 46.3, 25.0, 29.4)

data <- mdma_zuerich - mdma_basel

mean(data)

20.27

sd(data)

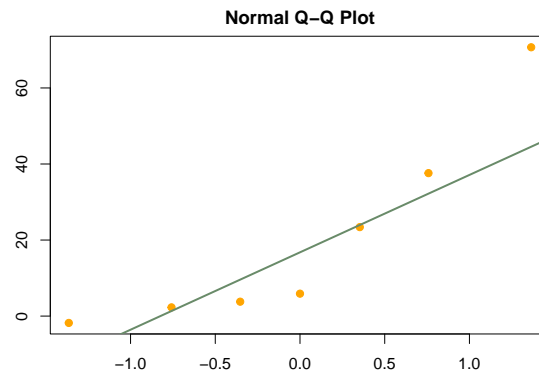
26.2723
```

At first glance, the situation seems clear: more MDMA is consumed in Zurich. However, we then realize that the differences show a rather large spread.

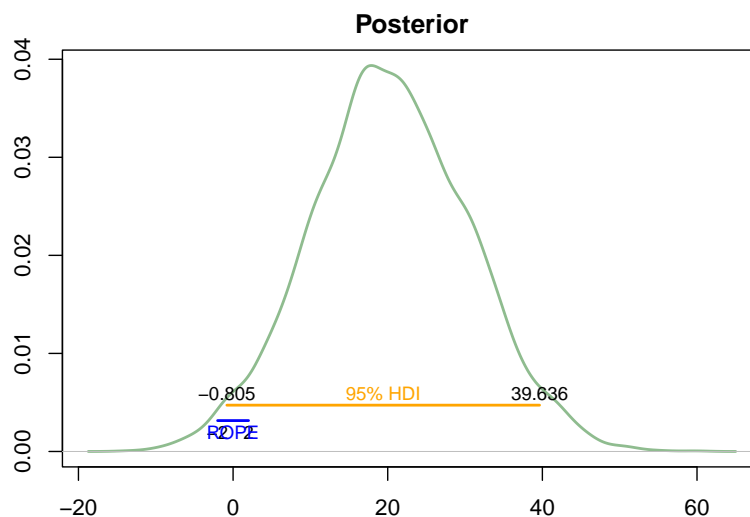
- b) If the day is defined as the experimental unit, then we can consider the samples to be paired. In this task, we will assume paired samples.

The cities of Zurich and Basel can be considered as different experimental units. In this case, we would consider the samples to be unpaired.

- c) At first glance, the QQ plot for the difference does not indicate a normal distribution. Since the data set is very small, a conclusive assessment is difficult.



- d) We choose a uniform distribution between -100 and 100 as the prior and choose the value from a) for the standard deviation.
- e) The mode of the posterior distribution is 20 and the HDI is $[-0.8, 39.6]$. We choose a ROPE around the zero value 0 because we want to test the null value that there is no difference between the two cities regarding MDMA use.



Technically, there is little overlap between HDI and ROPE, we cannot discard the null value 0 based on the data. However we choose the ROPE there is some overlap and the decision depends heavily on the width of the ROPE.

The data tend to indicate a relevant difference with regard to MDMA use in the two cities even though there is little overlap.

- f) However, if we now consider the two samples as unpaired, a completely different picture emerges.

```
data_z <- c(16.3, 12.7, 14.0, 53.3, 117, 62.6, 27.6)
data_b <- c(10.4, 8.91, 11.7, 29.9, 46.3, 25.0, 29.4)

sd(data_z)
sd(data_b)

par(mfrow=c(1,2))
qqnorm(data_z, pch=19, col="orange")
qqline(data_z, col="darkseagreen4", lwd=2)

qqnorm(data_b, pch=19, col="orange")
qqline(data_b, col="darkseagreen4", lwd=2)

modelString = "
data {
  int n;
  real y[n];
}

parameters {
  real<lower=50,upper=200> mu;
  real<lower=38.0,upper=38.1> sigma;
}

model {
  for (i in 1:n)
    y[i] ~ normal(mu,sigma);
}
" # close quote for modelString

model <- stan_model(model_code=modelString)

n      = length(data_z)
y      = data_z

stanFit = sampling( object=model , data=list( y = y , n = n ))

params_z = rstan::extract(stanFit)

modelString = "
data {
  int n;
```

```
real y[n];
}

parameters {
  real<lower=50,upper=200> mu;
  real<lower=13.6,upper=13.7> sigma;
}

model {
  for (i in 1:n)
    y[i] ~ normal(mu,sigma);
}
" # close quote for modelString

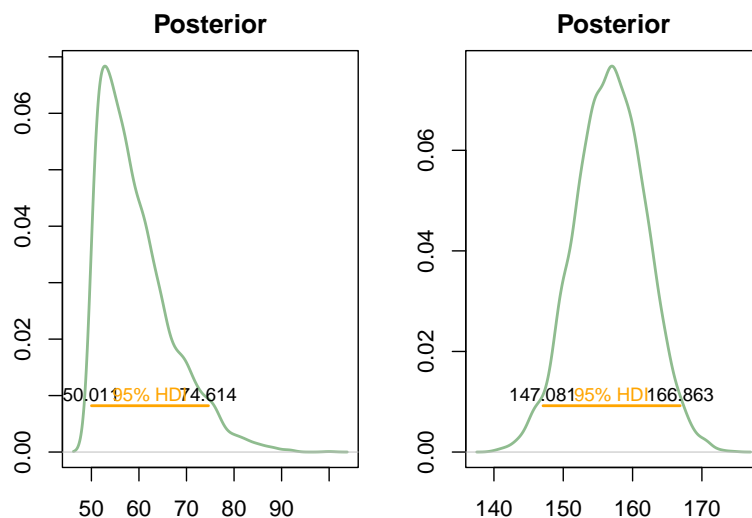
model <- stan_model(model_code=modelString)

n      = length(data_b)
y      = data_b

stanFit = sampling( object=model , data=list( y = y , n = n ))

params_c = rstan::extract(stanFit)

par(mfrow=c(1,2))
plot_posterior(params_z)
plot_posterior(params_b)
```



In this case, we would reject the decision because the two HDIs overlap considerably. This is because the spread of the MDMA's.

Solution 13.6

- a) For each pair of subjects (husband and wife), there are two measurements (height of the husband, height of the wife).

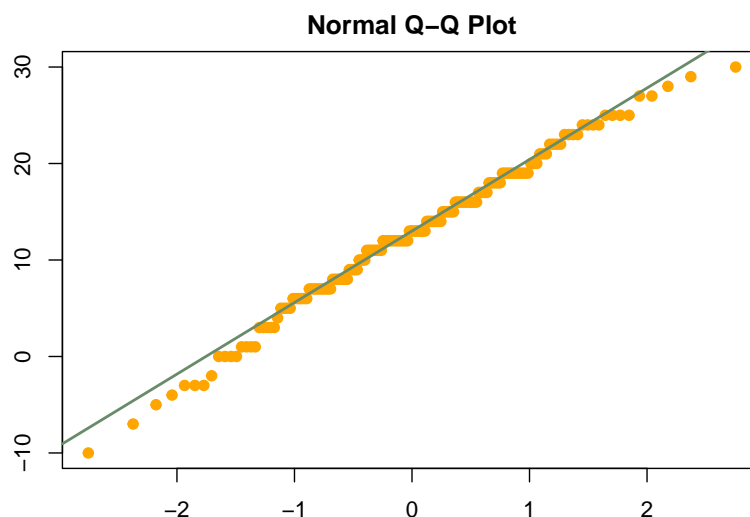
If the height difference of 13 cm is also valid within couples that would mean that tall men marry tall women and short men marry short women *on average*.

- b) The QQ plot clearly indicates a normal distribution.

```
data <- read.csv("husband_wife.csv")

data <- data[, "height.husband"] - data[, "height.wife"]

qqnorm(data, pch=19, col="orange")
qqline(data, col="darkseagreen4", lwd=2)
```



- c) We choose a uniform distribution between -50 and 50 as the prior distribution for the differences in height.

For the standard deviation

```
sd(data)

7.563503
```

The mode of the posterior distribution is 13 and the HDI is [11.4, 13.7]. We choose a ROPE around the zero value of 13 because we want to check whether the difference is approximately 13 cm. For a ROPE of [11, 15] we obtain

```
modelString = "
data {
  int n;
  real y[n];
}

parameters {
  real<lower=-50,upper=50> mu;
  real<lower=7.5,upper=7.6> sigma;
}

model {
  for (i in 1:n)
    y[i] ~ normal(mu,sigma);
}
" # close quote for modelString

model <- stan_model(model_code=modelString)

n      = length(data)
y      = data

stanFit = sampling( object=model , data=list( y = y , n = n ))

params = rstan::extract(stanFit)

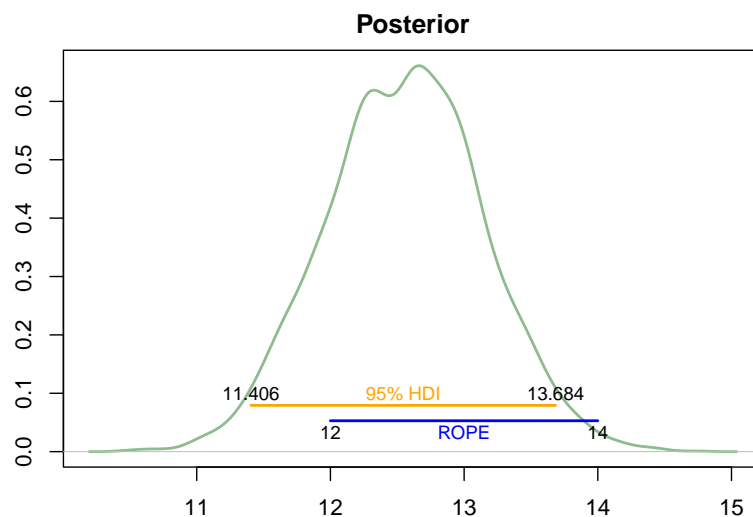
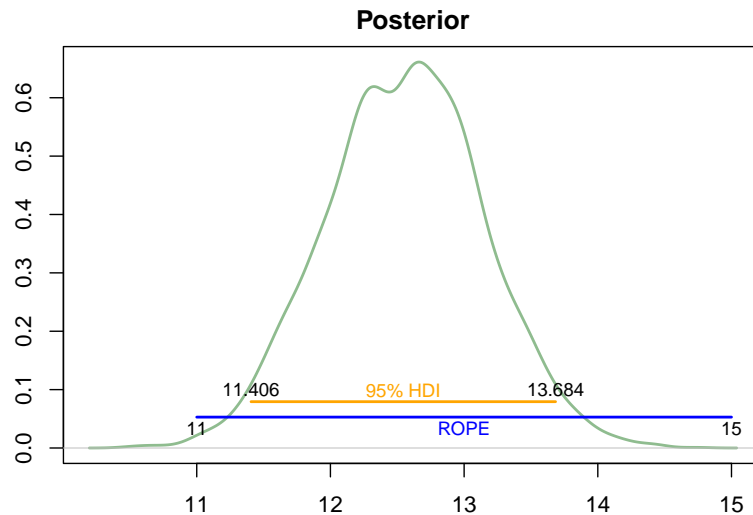
plot_posterior(params, Rope=c(11,15))
```

The HDI is included in ROPE, and thus the data statistically confirm the height difference of 13 cm within the couples. We conclude that tall men marry *on average* indeed tall women and short men marry short women.

However; if we choose a narrower ROPE, say [12, 14], the situation is not so clear anymore.

Solution 13.7

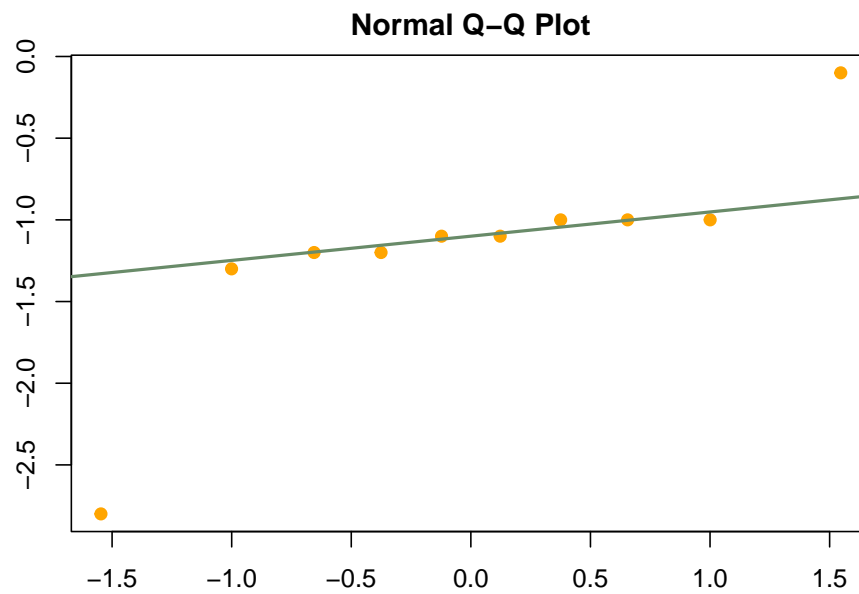
- a) It is a paired test because two measurements were taken on one unit of investigation (patient).



- b) The QQ plot of the differences does indicate a normal distribution, except the tails. Since the data set is very small, a conclusive assessment is difficult. However, the data set can also be part of an approximated normal distribution curve.

```
df_1 <- c(39.1, 39.3, 38.9, 40.6, 39.5, 38.4, 38.6, 39.0, 38.6, 39.2)
df_2 <- c(38.1, 38.3, 38.8, 37.8, 38.2, 37.3, 37.6, 37.8, 37.4, 38.1)
data <- df_2 - df_1

qqnorm(data, pch=19, col="orange")
qqline(data, col="darkseagreen4", lwd=2)
```



- c) For the prior distribution, we choose a uniform distribution between -10 and 10 .

For the standard deviation

```
sd(data)
```

```
0.6596295
```

```
modelString = "  
data {  
  int n;  
  real y[n];  
}  
  
parameters {  
  real<lower=-10,upper=10> mu;  
  real<lower=0.6,upper=0.7> sigma;  
}  
  
model {  
  for (i in 1:n)  
    y[i] ~ normal(mu,sigma);  
}
```



```
" # close quote for modelString

model <- stan_model(model_code=modelString)

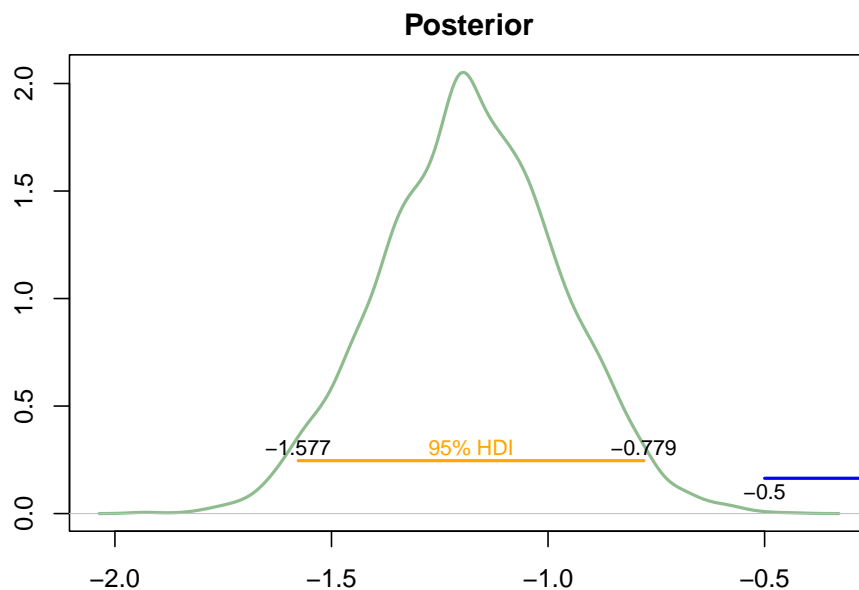
n      = length(data)
y      = data

stanFit = sampling( object=model , data=list( y = y , n = n ))

params = rstan::extract(stanFit)

plot_posterior(params, Rope=c(-0.5,0.5))
```

- d) The mode of the posterior distribution is -1.2 and the HDI is $[-1.6, -0.78]$. We choose the ROPE around the zero value 0 because we want to check whether the difference shows a statistically relevant deviation from 0 .
- e) We consider a deviation of up to 0.5 from the zero value 0 to be equivalent to the zero value, namely no effect



- f) The HDI and ROPE do not overlap, and thus the data confirm a statistically significant antipyretic effect of the drug.