

**⚠️ RANDOM NUMBER DIFFERENCES**

The default random number generator (RNG) was changed in R version 3.6.0. The solutions to problems in this course are based on the newer RNG settings. Therefore, if you're using R older than 3.6.0, your answers will differ from the solutions for problems that require sampling random observations (these problems usually involve setting the RNG seed with "set.seed").

In some cases, you may still run into this problem while using newer version of R. This can happen if you load an older R session from a .RData file where the older RNG settings were used (loading the local .RData file is the default behavior of RStudio).

You can verify your RNG settings with the following command:

```
RNGkind()
```

This should return 3 values: "Mersenne-Twister", "Inversion", "Rejection". If you see that the last value is "Rounding", you are using the older settings (this is not what you want!). To fix the issue, run the following command.

```
RNGkind("Mersenne-Twister", "Inversion", "Rejection")
```

## Hierarchical Clustering Exercises #1

1/1 point (graded)

Create a random matrix with no correlation in the following way:

```
set.seed(1)
m = 10000
n = 24
x = matrix(rnorm(m*n),m,n)
colnames(x)=1:n
```

Run hierarchical clustering on this data with the `hclust()` function with default parameters to cluster the columns. Create a dendrogram.

From the dendrogram which pairs of samples are the furthest away from each other?

☐ 7 and 23

☐ 19 and 14

☐ 1 and 16

☒ 17 and 9




Explanation

```
hc = hclust( dist( t(x)))  
plot(hc)
```

Submit

You have used 2 of 2 attempts

 Answers are displayed within the problem

## Hierarchical Clustering Exercises #2

1/1 point (graded)

Set the seed at 1 with `set.seed(1)` and replicate the creation of this matrix 100 times:

```
m = 10000  
n = 24  
x = matrix(rnorm(m*n),m,n)
```

then perform hierarchical clustering as in the solution to question 2.4.1 and find the number of clusters if you use `cutree()` at height 143. Note that this number is a random variable.

Based on the Monte Carlo simulation, what is the population standard deviation of this random variable?

0.8986657

✓ Answer: 0.8986657

**0.8986657**

Explanation

```
set.seed(1)  
m = 10000  
n = 24  
nc = replicate(100,{  
  x = matrix(rnorm(m*n),m,n)  
  hc = hclust( dist( t(x)))  
  length(unique(cutree(hc,h=143)))  
})  
plot(table(nc)) ## look at the distribution  
popstd(nc)
```

Submit

You have used 1 of 2 attempts

 Answers are displayed within the problem