Homework due Nov 13, 2023 10:13 CST

For the following questions, use the data loaded with:

```
library(tissuesGeneExpression)
data(tissuesGeneExpression)
```

---

## MDS Exercises #1

1/1 point (graded)

In these exercise we will demonstrate the relantionship between the SVD and the output of `cmdscale()`, the function in R that performs MDS.

Using the `z` we computed in SVD Exercises #4:

```
y = e - rowMeans(e)
s = svd(y)
z = s$d * t(s$v)
```

we can make an MDS plot:

```
library(rafalib)
ftissue = factor(tissue)
mypar(1,1)
plot(z[1,],z[2,],col=as.numeric(ftissue))
legend("topleft",levels(ftissue),col=seq_along(ftissue),pch=1)
```

Now run the function `cmdscale()` on the original data:

```
d = dist(t(e))
mds = cmdscale(d)
```

What is the correlation between the first row of `z` and the first column in `mds`?

-1 ✔ Answer: -1

$-1$

Explanation

```
cor(z[1,],mds[,1])
```

Submit You have used 1 of 5 attempts

Answers are displayed within the problem

## MDS Exercises #2

1/1 point (graded)

What is the correlation between the second row of `z` and the second column of `mds`?

-1 ✔ Answer: -1

$-1$

Explanation

```
cor(z[2,],mds[,2])
```

Submit You have used 1 of 5 attempts

Answers are displayed within the problem

## MDS Exercises #3

1/1 point (graded)

Note that the MDS plot is not the same:

```
library(rafalib)
ftissue = factor(tissue)
mypar(1,2)
plot(z[1,],z[2,],col=as.numeric(ftissue))
legend("topleft",levels(ftissue),col=seq_along(ftissue),pch=1)
plot(mds[,1],mds[,2],col=as.numeric(ftissue))
```

Given the answer to MDS Exercises #1 and #2, what do we have to do to `z[1,]` and `z[2,]` to get a practically identical plot?

- ○ It is impossible
- ○ Use the columns instead `z[,1]` and `z[,2]`
- ○ Remove the row means from `e` before computing the distance
- ◉ multiply `z[1,]` and `z[2,]` by -1

✔

Explanation
From the answer to the previous question we see that we have a flipped sign in both dimensions.

Submit You have used 1 of 2 attempts

ℹ Answers are displayed within the problem

## MDS Exercises #4

1/1 point (graded)
Load the following dataset:

```
library(GSE5859Subset)
data(GSE5859Subset)
```

Compute the SVD and compute `z`:

```
s = svd(geneExpression-rowMeans(geneExpression))
z = s$d * t(s$v)
```

Which dimension of `z` most correlates with the outcome `sampleInfo$group`?

1 ✔ Answer: 1

1

Explanation

```
which.max(cor(sampleInfo$g,t(z)))
```

Submit You have used 1 of 5 attempts

Answers are displayed within the problem

## MDS Exercises #5

1/1 point (graded)

Continue working with the `z` calculated from the **GSE5859Subset** data.

What is this max correlation?

0.6236858 ✔ Answer: 0.6236858

0.6236858

Explanation

```
max(cor(sampleInfo$g,t(z)))
```

Submit You have used 1 of 5 attempts

Answers are displayed within the problem

## MDS Exercises #6

1/1 point (graded)
Continue working with the `z` calculated from the **GSE5859Subset** data.

Which dimension of `z` has the *second highest* correlation with the outcome `sampleInfo$group`?

6 ✔ Answer: 6

**6**

Explanation

```
which.max(cor(sampleInfo$g,t(z))[-1]) + 1
```

We add 1 because we took out the first.

Submit You have used 1 of 5 attempts

Answers are displayed within the problem

## MDS Exercises #7

2/2 points (graded)
Note these measurements were made during two months:

```
sampleInfo$date
```

We can extract the month this way:

```
month = format( sampleInfo$date, "%m")
month = factor( month)
```

Which dimension of `z` has the highest correlation with the outcome `month`?

1 ✔ Answer: 1

**1**

Explanation

```
which.max(cor( as.numeric(month), t(z)))
```

What is this correlation?

.8297915 ✔ Answer: 0.8297915

**.8297915**

Explanation

```
max(cor( as.numeric(month), t(z)))
```

Note that the same dimension is correlated with both the group and the date. Note also that these are correlated:

```
table(sampleInfo$g,month)
```

So is this first dimension related directly to group or is it related only through the month? Note that the correlation with month is higher. This is related to *batch effects* which we will learn about later.

Submit You have used 1 of 5 attempts

Answers are displayed within the problem

## MDS Exercises #8 (ADVANCED)

1/1 point (graded)

Note: this is an advanced question. Please feel free to research this question online.

In MDS Exercises #7 we saw that that one of the dimensions was highly correlated to the `sampleInfo$group`. Now take the 5th column of $\mathbf{U}$ and stratify by the gene chromosome. Remove `chrUn` and make a boxplot of the values of $\mathbf{U_6}$ stratified by chromosome.

Which chromosome looks different from the rest?

Copy and paste the name as it appears in `geneAnnotation`.

chrY ✔ Answer: chrY

Explanation

```
result = split(s$u[,6],geneAnnotation$CHR)
result = result[ which(names(result)!="chrUn") ]
boxplot(result,range=0)
boxplot(result,range=0,ylim=c(-0.025,0.025))
medians = sapply(result,median)
names(result)[ which.max(abs(medians)) ]
```

Given the answer to this question, any guesses as to what `sampleInfo$group` represents?

Submit You have used 1 of 10 attempts

Answers are displayed within the problem