

[Course](#) > [Section 6: Model Fit...](#) > [6.1: Case Study: M...](#) > [Comprehension Ch...](#)

Comprehension Check: Dimension Reduction

The [dimension reduction section of the textbook](#) may help with some of these exercises.

Q1

1/1 point (graded)

We want to explore the `tissue_gene_expression` predictors by plotting them.

```
data("tissue_gene_expression")  
dim(tissue_gene_expression$x)
```

We want to get an idea of which observations are close to each other, but, as you can see from the dimensions, the predictors are 500-dimensional, making plotting difficult. Plot the first two principal components with color representing tissue type.

Which tissue is in a cluster by itself?

☐ cerebellum

☐ colon

☐ endometrium

☐ hippocampus

☐ kidney

☒ liver

☐ placenta



Explanation

The plot can be made using the following code:

```
pc <- prcomp(tissue_gene_expression$x)
data.frame(pc_1 = pc$x[,1], pc_2 = pc$x[,2],
            tissue = tissue_gene_expression$y) %>%
  ggplot(aes(pc_1, pc_2, color = tissue)) +
  geom_point()
```

We can see that liver clusters alone in the lower right-hand corner of the plot.

Submit

You have used 1 of 3 attempts

i Answers are displayed within the problem

Q2

1/1 point (graded)

The predictors for each observation are measured using the same device and experimental procedure. This introduces biases that can affect all the predictors from one observation. For each observation, compute the average across all predictors, and then plot this against the first PC with color representing tissue. Report the correlation.

What is the correlation?

0.597

✓ Answer: 0.597

0.597

Explanation

The plot and correlation can be generated using the following code:

```
avgs <- rowMeans(tissue_gene_expression$x)
data.frame(pc_1 = pc$x[,1], avg = avgs,
            tissue = tissue_gene_expression$y) %>%
  ggplot(aes(avgs, pc_1, color = tissue)) +
  geom_point()
cor(avgs, pc$x[,1])
```

Submit

You have used 1 of 10 attempts

i Answers are displayed within the problem

Q3

1/1 point (graded)

We see an association with the first PC and the observation averages. Redo the PCA but only after removing the center. Part of the code is provided for you.

```
#BLANK
pc <- prcomp(x)
data.frame(pc_1 = pc$x[,1], pc_2 = pc$x[,2],
           tissue = tissue_gene_expression$y) %>%
  ggplot(aes(pc_1, pc_2, color = tissue)) +
  geom_point()
```

Which line of code should be used to replace #BLANK in the code block above?

- ☐ `x <- with(tissue_gene_expression, sweep(x, 1, mean(x)))`
- ☐ `x <- sweep(x, 1, rowMeans(tissue_gene_expression$x))`
- ☐ `x <- tissue_gene_expression$x - mean(tissue_gene_expression$x)`
- ☒ `x <- with(tissue_gene_expression, sweep(x, 1, rowMeans(x)))`



Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Q4

1/1 point (graded)

For the first 10 PCs, make a boxplot showing the values for each tissue.

For the 7th PC, which two tissues have the greatest median difference?

Select the TWO tissues that have the greatest median difference.

☐ cerebellum

☒ colon

☐ endometrium

☐ hippocampus

☐ kidney

☐ liver

☒ placenta



Explanation

The boxplots for the first 10 PCs can be made using this code:

```
for(i in 1:10){  
  boxplot(pc$x[,i] ~ tissue_gene_expression$y, main = paste("PC", i))  
}
```

Submit

You have used 1 of 3 attempts

i Answers are displayed within the problem

Q5

1/1 point (graded)

Plot the percent variance explained by PC number. Hint: use the `summary` function.

How many PCs are required to reach a cumulative percent variance explained greater than 50%?

3

✓ Answer: 3

3

Explanation

The plot can be made using the following code: `plot(summary(pc)$importance[3,])`.

Submit

You have used 1 of 10 attempts

i Answers are displayed within the problem

Ask your questions or make your comments about Dimension Reduction here! **Remember, one of the best ways to reinforce your own learning is by explaining something to someone else, so we encourage you to answer each other's questions (without giving away the answers, of course).**

Some reminders:

- Search the discussion board before posting to see if someone else has asked the same thing before asking a new question.
- Please be specific in the title and body of your post regarding which question you're asking about to facilitate answering your question.
- Posting snippets of code is okay, but posting full code solutions is not.
- If you do post snippets of code, please format it as code for readability. If you're not sure how to do this, there are instructions in a pinned post in the "general" discussion forum.

Discussion: Dimension Reduction

[Show Discussion](#)

Topic: Section 6: Model fitting and recommendation systems / 6.1.2: Dimension Reduction