Statistical Analysis in R(Session 4) - Exercises and Solutions

Exercise 1: Compute Summary Statistics

Using the `mtcars` dataset, calculate the following:

1. Mean, median, and standard deviation of `mpg`.
2. Minimum and maximum values of `hp`.

Solution:

```
mean(mtcars$mpg)
median(mtcars$mpg)
sd(mtcars$mpg)
min(mtcars$hp)
max(mtcars$hp)
```

Exercise 2: Compute Correlations

1. Compute the Pearson correlation between `mpg` and `hp`.
2. Create a correlation matrix for all numeric variables.

Bonus: Visualize as a heatmap

Solution:

```
cor(mtcars$mpg, mtcars$hp)
cor(mtcars)
```

Exercise 3: Perform a T-test

1. Test if the mean `mpg` differs between automatic and manual transmission (`am`).

Solution:

```
t.test(mpg ~ am, data = mtcars)
```

Exercise 4: Compute a Confidence Interval

1. Compute a 95% confidence interval for the mean `mpg`.

Solution:

```
mean_mpg <- mean(mtcars$mpg)
se_mpg <- sd(mtcars$mpg) / sqrt(nrow(mtcars))
ci <- mean_mpg + c(-1.96, 1.96) * se_mpg
ci
```

Exercise 5: One-way ANOVA

1. Test if there are significant differences in `mpg` across different numbers of cylinders (`cyl`).

Solution:

```
anova_result <- aov(mpg ~ factor(cyl), data = mtcars)
summary(anova_result)
```

Exercise 6: Perform Principal Component Analysis(PCA)

1. Conduct PCA on the `mtcars` dataset (excluding categorical variables) and visualize the results using a biplot.
2. Visualize eigenvalues
3. Learn about PCA: https://rpubs.com/KarolinaSzczesna/862710

Solution:

```
library(factoextra)
mtcars_pca <- prcomp(mtcars[, sapply(mtcars, is.numeric)], scale. = TRUE)
biplot(mtcars_pca)
fviz_eig(mtcars_pca)
```

Extra:

1. Perform differential expression analysis by using DESeq2 and Gene Expression data

Solution:

https://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html