

Charlee Cobb - Transcriptomics, Exercise 4

netID: cac9995

Charlee Cobb

2023-02-10

M4_1.1) In what scenarios is it better to use ANOVA compared to t-test? More specifically when can you NOT use a t-test? It's better to use ANOVA when you are comparing more than two samples.

M4_1.2) How is the ANOVA statistics calculated? ANOVA calculates the variation within the groups, and then compares that to variation between groups.

M4_1.3) What assumptions does ANOVA make about the data? ANOVA assumes that the data is normally distributed, and that the data is independent.

M4_2.1) What are some differences between pearson and spearman correlation? The pearson correlation compares two variable's scores that are normally distributed and have a linear relationship. Spearman works on non-linear relationships, and focuses on ranking data.

M4_2.2) What is the range of values one can obtain from a correlation analysis? The range of values falls between 1 and -1.

M4_3.1) What are the two variables that are needed to draw a straight line? You need a Y and an x variables, the response and explanatory variables.

M4_3.2) What command can we use to create a linear model? the `lm()` function is used to make a linear model.

M4_3.3) What command do we use to predict the values using the linear model? `predict()` function is used to predict the values.

M4_4.1) What is a limitation of the CHI-Square test? Values are not as dependable when proving the null hypothesis that there is no association.

M4_4.2) What is the null hypothesis of a fisher's exact test? A random table will have a probability equal to or less than input table's probability.

M4_4.3) Give an example of Hypergeometric tests can be used with biological data. You can use Hypergeometric tests to get the probability of sampling genes with a specific functionality in a target cancer from a database.

M4_5.1) Why is it important to correct for multiple hypothesis testing? It's important to correct for multiple hypothesis testing because you need to check for false positive and false negative conclusions made on the null hypothesis.

M4_5.2) Of the two methods discussed (Bonferroni and FDR) which method is more appropriate for life science (when we have few replicates). FDR method is better for life science

M4_5.3) What is the difference between a TypeI and TypeII error? Type one is false positive errors, and type two is false negative. Type one falsely rejects the null hypothesis, and type two falsely accepts the null hypothesis.

M4_6.1) Install the reshape2 package in R. What does the melt function do? The melt() function creates one single data frame by combining separate vectors. Unlike cbind() or rbind(), you can change the shape of the data frame by making it longer or wider.

M4_6.2) Check out the help documentation for rnorm using help("rnorm"). What are the first three arguments of the function? the first three arguments of the function are a vector, mean (vector of means), then the standard deviation (a vector of standard deviations).

```
help("rnorm")
```

M4_6.3) How is the slope of the linear model and the difference in the means related in this example? The difference is represented through the R-squared mode. Are the difference in the means the residuals?

M4_7.1) What is the difference between the lm() function and aov() function? Do they give you the same result when looking at just one factor with two levels? They function the same way when looking at one factor with two levels. lm() looks for the slope made in the data, and looks at each slope separately. aov() looks at variation between groups.

M4_8.1) What is the argument for TukeyHSD and what analysis does it perform? TukeyHSD() takes aov() results and performs a pairwise comparison with all combos.