Descriptive Statistics and Statistical Tests

In this exercise, you will use R to calculate descriptive statistics, conduct a t-test, assess variable correlation, and compare group means. Create code to obtain the result and/or answer the following questions. Submit your code and the answers in an R file (.R) or text file (.txt). Make sure to include your answers as comments.

The following data sets have been provided:

- **matts_movies.csv**: movie data set created by my brother ("Movie.Name" = Name of movie, "Director" = Director of movie, "Release.Year" = Year the movie was released, "My.Rating" = My brother’s rating of the movie, "Genre" = The movie genre, "Own" = Whether my brother owns a copy of the movie)
- **mine_classification_with_lidar.csv**: a data set of land cover types with associated variables
- **wetland_binary.csv**: a data set of wetland and upland examples with associated landscape variables

**T1**: How many different land cover types ("class") are differentiated in the mine classification data set? (2 Points)

**T2**: Which land cover type ("class") has the highest mean normalized difference vegetation index ("ndvi")? (2 Points)

**T3**: Which land cover type ("class") has the highest mean height ("diff")? (2 Points)

**T4**: What is the standard deviation of NDVI ("ndvi") for the forest class ("class" forest)? (2 Points)

**T5**: What is the mean rating ("My.Rating") for the dramas ("Genre" "Drama") in the movie data set? (2 Points)

**T6**: Which genre ("Genre") has the largest range of ratings ("My.Rating") in the movies data set? (2 Points)

**T7**: Using the Spearman method, what is the correlation coefficient between slope in degrees ("slp_d") and topographic dissection ("diss_a") in the wetland data set? (2 Points)

**T8**: Use a t-test to assess if average slope ("slp_d") is different between wetlands ("class" wet) and not wetlands ("class" not). What does the result suggest? (2 Points)

**T9**: Use a Mann-Whitney U test to assess if average slope ("slp_d") is different between wetlands ("class" wet) and not wetlands ("class" not). What does the result suggest? (2 Points)

**T10**: Use a t-test to assess if the average rating ("My.Rating") is different between dramas ("Genre" Drama) and comedies ("Genre" Comedy). What does the result suggest? (2 Points)

**T11**: Use a Mann-Whitney U test to assess if the average rating ("My.Rating") is different between dramas ("Genre" Drama) and comedies ("Genre" Comedy). What does the result suggest? (2 Points)

**T12**: Use ANOVA to assess whether there is statistical difference in mean NDVI ("ndvi") between at least two land cover type ("class"). What does the result suggest? (4 Points)
T13: Use Tukey’s Honest Significant Difference test to assess whether there is a difference between the forest (“class” forest) and shrub (“class” shrub) land cover types specifically. What does the result suggest? (2 Points)

T14: Create a QQ plot for the model residuals to assess whether the residuals are normally distributed. Does the plot suggest that there is an issue with this assumption? (2 Points)

T15: Use the Bartlett Test of Homogeneity of Variance to assess whether there is consistent variance between the classes. What does this test suggest? (2 Points)

T16: Use the Bonferroni Outlier Test to test for outliers. Does this test suggest the presence of outliers?

T17: Use the Kruskal-Wallis Rank Sum Test to assess whether at least two cover types ("class") have a different mean NDVI ("ndvi"). What does this test suggest? (4 Points)

T18: Use a pairwise Kruskal-Wallace test to assess whether the herbaceous (“class” herb) and forest (“class” forest) types have different mean NDVI at a 95% confidence interval. What does this test suggest? (2 Points)