Case Study 10: Are hit songs getting longer?

Situation

You are tasked to discover if current top songs in the United States are longer then the top songs of the 1990s. Being able to detect this trend is crucial for song writers,and they need your help to do it. Below are two links that list out the top songs from the 90s and 2000s:

2000: https://en.wikipedia.org/wiki/List_of_Billboard_Hot_100_number-one_singles_of_the_2000s

1990: https://en.wikipedia.org/wiki/List_of_Billboard_Hot_100_number-one_singles_of_the_1990s

Learning Objectives

- Understand how to carry out a random sample, and the impacts this type of sampling has on scope of inference

- Identify and justify the appropriate research methods to answer your research question

- Write an appropriate scope of inference for this study. Understand the two questions that make up a scope of inference.

Questions

- Explore each website. Detail how you will select a random number of songs from each list. Hint: Songs need to have an equal probability of being selected. Simply "choosing them at random" by yourself is not a random sample. Post your plan on the D2L discussion board "Random Selection of Songs."

- Using your detailed plan above, randomly sample 10 songs from each website (Year). Do some additional research and look up the duration of each of the randomly selected songs. Download the excel file titled "SongData." You will see two columns: *Duration* and *Year*. For each of the 10 selected songs in the 1990s, input the duration in the *Duration* column next to a 1990 input in the *Year* column. Repeat this process for the other 10 songs in the 2000s.

- Now that you have built your data set, upload your data into R. For review, please see the video ImportingDataRStudio on D2L. Make sure that your data set name is SongData for the following code to work. You will see the name of your data set show up in the top right panel and no error codes below if you successfully have uploaded your data set.

- Plot your data using the given code. Describe characteristics of these data for each group. Which group has the higher mean / median duration?

- Identify your method of analysis. Justify and explain what type of procedure you will use to analyze these data.

- Check your assumptions. Is there any evidence that one (or more) of your assumptions may be violated? Justify your answer.

- Carry out your test. Regardless of your answer above, carry out a theoretical procedure to answer your research question. Report the p-value, test statistic, and the distribution we assume our statistic follows below.

- Write an appropriate conclusion below. Post your conclusion to the discussion post "Are songs getting longer?" on D2L. Include your scope of inference. Be sure to include information from the previous question.

- Reflect on the learning objectives above. If you have any questions or comments, please post on the "Song Study Questions" discussion post.