UNIVERSITY OF STRATHCLYDE
DEPARTMENT OF MATHEMATICS & STATISTICS

MM104: Statistics and Data Presentation Semester 2
MM107: Statistics and Data Presentation

# Project 6: Correlation and Linear Regression

## Overview

In this project you will

1. Investigate the relationship between numerical variables.
2. Use Minitab to determine whether or not two variables are correlated and if so fit a linear regression line to the data.
3. Interpret the outputs and estimates of the fitted linear regression model and validate the assumptions of fitted linear model.
4. Use your model to make a prediction and compare this to the real world result.
5. Use Word and PowerPoint for reports and presentations.
6. Further develop your presentation skills.

## The Data Set

The data-set comes from a Spotify playlist, the MM104 Playlist (can be found here if interested `https://open.spotify.com/playlist/2NVgq429BuAJURKbILu3FH?si=SOCrD8QoQtSxVvVR6hmFnQ`). Ainsley asked all the members of staff and tutors involved in MM104 to provide her with songs (these songs could be your favourite songs, or songs that mean something to the person e.g. first dance at their wedding). Ainsley also asked friends of MM104, to give her their favourite songs too (in order to get the sample size up). These songs were then turned into a playlist on Spotify, and then from this playlist various information could be extracted. Each participant in the study gave approximately 5 songs to the data set.

The playlist is made up a collection of songs from a variety of artists, genre and not all songs are performed in English. Note that some songs contain explicit language these are clearly marked on the playlist.

| Variable Name | Description and Research Question |
|---|---|
| **Title** | Song Title |
| **Artist** | Band/Singer |
| **Release** | Year song was released |
| **Genre** | Genre of music that the song best aligns with. |
| **Length** | Length of the song in minutes and seconds. |
| **Contributor** | Classification of the contributor: Maths and Stats Staff Member, PhD student, Other (where Other includes: undergraduate tutors, ex-colleagues, Ainsley's parents etc.) |
| **BPM** | Number of beats per minute a song has. |
| **Energy** | The energy of the song, the higher the value the more energetic the song (score ranges from 0 to 100). |
| **Dance** | The dance-ability of the song, the higher the value, the easier it is to dance to this song (score ranges from 0 to 100). |
| **Happiness** | The happiness of the song, the higher the value, the more positive the mood of the song (score ranges from 0 to 100). |
| **Acoustic** | How acoustic a song is, the higher the score the more acoustic (score ranges from 0 to 100). |
| **Popularity** | How popular the song is, the higher the value the more popular (score ranges from 0 to 100). |

Each group will work with one of the data sets as shown below. each group is interested in finding out if a particular variable is attributed with a songs' popularity.

| Group Number | Topics |
|---|---|
| 1 | Popularity and BPM |
| 2 | Popularity and Energy |
| 3 | Popularity and Dance |
| 4 | Popularity and Happiness |
| 5 | Popularity and Acoustic. |

# Tasks

This week you are expected to at least:

- Conduct background research on the variables you have been assigned.
- Produce a scatter plot.
- Calculate the correlation coefficient.
- If appropriate, perform a linear regression. To perform a linear regression the correlation coefficient must have an absolute value of greater than 0.4.
- Comment on the assumptions of the linear regression.

# Report and Presentation

Your Word report should be at least 3 to 4 pages, and at least 750 words (excluding title page and references). Your report should have sections which correspond to the sections in your PowerPoint presentation as well as a references section. Since there is quite a lot involved in a linear regression, the time limit of your presentation has been extended to between 6 and 7 minutes.

Remember: the first or second page should contain a declaration section. This should list the names of all group members and sentence outlining each person's contribution to the project. Each person should sign this sheet electronically.