Appendix VII: Model Summaries

In [1]:
```
"""Imports necessary packages"""

import itertools
import math
from typing import Dict, Iterable, List, Union

import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import pylab
import scipy
import scipy.stats as stats
import seaborn as sns
import statsmodels.api as sm
from sklearn.model_selection import train_test_split

sns.set_style("whitegrid")
```

In [2]:
```
data = pd.read_csv("D:/School/frequentist-statistics/ITM-song-popularity/database/it
data = data.drop("Unnamed: 0", axis=1)
```

In [3]:
```
models = ["popularity_abs ~ age_days + complexity + track_number", "popularity_norm
```

In [4]:
```
for model_str in models:
    print("--- SUMMARY FOR %s ---" % model_str)
    model = sm.formula.ols(formula=model_str, data=data)
    model_fitted = model.fit()
    print(model_fitted.summary())
```

```
--- SUMMARY FOR popularity_abs ~ age_days + complexity + track_number ---
                            OLS Regression Results
==============================================================================
Dep. Variable:         popularity_abs   R-squared:                       0.562
Model:                            OLS   Adj. R-squared:                  0.546
Method:                 Least Squares   F-statistic:                     35.02
Date:                Sun, 06 Jun 2021   Prob (F-statistic):           1.14e-14
Time:                        00:28:33   Log-Likelihood:                -280.76
No. Observations:                  86   AIC:                             569.5
Df Residuals:                      82   BIC:                             579.3
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept       39.9574      2.803     14.257      0.000      34.382      45.533
age_days        -0.0043      0.001     -7.169      0.000      -0.006      -0.003
complexity      26.3952      3.886      6.792      0.000      18.665      34.126
track_number    -0.9262      0.199     -4.648      0.000      -1.323      -0.530
==============================================================================
Omnibus:                        6.298   Durbin-Watson:                   1.716
Prob(Omnibus):                  0.043   Jarque-Bera (JB):                5.054
Skew:                           0.486   Prob(JB):                       0.0799
Kurtosis:                       2.318   Cond. No.                     1.72e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly spe
```

```
cified.
[2] The condition number is large, 1.72e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
--- SUMMARY FOR popularity_norm ~ age_days + complexity + track_number ---
                            OLS Regression Results
==============================================================================
Dep. Variable:        popularity_norm   R-squared:                       0.562
Model:                            OLS   Adj. R-squared:                  0.546
Method:                 Least Squares   F-statistic:                     35.02
Date:                Sun, 06 Jun 2021   Prob (F-statistic):           1.14e-14
Time:                        00:28:33   Log-Likelihood:                 38.609
No. Observations:                  86   AIC:                            -69.22
Df Residuals:                      82   BIC:                            -59.40
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.4624      0.068      6.764      0.000       0.326       0.598
age_days      -0.0001   1.48e-05     -7.169      0.000      -0.000   -7.66e-05
complexity     0.6438      0.095      6.792      0.000       0.455       0.832
track_number  -0.0226      0.005     -4.648      0.000      -0.032      -0.013
==============================================================================
Omnibus:                        6.298   Durbin-Watson:                   1.716
Prob(Omnibus):                  0.043   Jarque-Bera (JB):                5.054
Skew:                           0.486   Prob(JB):                       0.0799
Kurtosis:                       2.318   Cond. No.                     1.72e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly spe
cified.
[2] The condition number is large, 1.72e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
--- SUMMARY FOR popularity_abs ~ age_days + complexity + track_number + track_number
*duration + danceability + duration ---
                            OLS Regression Results
==============================================================================
Dep. Variable:         popularity_abs   R-squared:                       0.613
Model:                            OLS   Adj. R-squared:                  0.583
Method:                 Least Squares   F-statistic:                     20.82
Date:                Sun, 06 Jun 2021   Prob (F-statistic):           1.77e-14
Time:                        00:28:33   Log-Likelihood:                -275.45
No. Observations:                  86   AIC:                             564.9
Df Residuals:                      79   BIC:                             582.1
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
=====
                             coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
-----
Intercept                30.5807      3.977      7.689      0.000      22.665      3
8.497
age_days                 -0.0048      0.001     -8.007      0.000      -0.006      -
0.004
complexity                6.1654     10.079      0.612      0.542     -13.896      2
6.227
track_number              0.2650      0.557      0.476      0.636      -0.844
1.374
duration                 28.9420     11.687      2.477      0.015       5.680      5
2.204
track_number:duration    -2.1057      0.907     -2.323      0.023      -3.910      -
0.301
danceability             15.5539      7.464      2.084      0.040       0.697      3
0.411
==============================================================================
Omnibus:                        3.772   Durbin-Watson:                   1.569
```

```
Prob(Omnibus):                  0.152   Jarque-Bera (JB):                 3.679
Skew:                           0.458   Prob(JB):                         0.159
Kurtosis:                       2.568   Cond. No.                      6.31e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly spe
cified.
[2] The condition number is large, 6.31e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
--- SUMMARY FOR popularity_norm ~ age_days + complexity + track_number + track_numbe
r*duration + danceability + duration ---
                            OLS Regression Results
==============================================================================
Dep. Variable:        popularity_norm   R-squared:                       0.613
Model:                            OLS   Adj. R-squared:                  0.583
Method:                 Least Squares   F-statistic:                     20.82
Date:                Sun, 06 Jun 2021   Prob (F-statistic):           1.77e-14
Time:                        00:28:33   Log-Likelihood:                 43.915
No. Observations:                  86   AIC:                            -73.83
Df Residuals:                      79   BIC:                            -56.65
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                             coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
Intercept                  0.2337      0.097      2.409      0.018       0.041
0.427
age_days                  -0.0001   1.46e-05     -8.007      0.000      -0.000   -8.7
9e-05
complexity                 0.1504      0.246      0.612      0.542      -0.339
0.640
track_number               0.0065      0.014      0.476      0.636      -0.021
0.034
duration                   0.7059      0.285      2.477      0.015       0.139
1.273
track_number:duration     -0.0514      0.022     -2.323      0.023      -0.095      -
0.007
danceability               0.3794      0.182      2.084      0.040       0.017
0.742
==============================================================================
Omnibus:                        3.772   Durbin-Watson:                   1.569
Prob(Omnibus):                  0.152   Jarque-Bera (JB):                3.679
Skew:                           0.458   Prob(JB):                        0.159
Kurtosis:                       2.568   Cond. No.                     6.31e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly spe
cified.
[2] The condition number is large, 6.31e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
--- SUMMARY FOR popularity_abs ~ track_number + duration + danceability + age_days -
--
                            OLS Regression Results
==============================================================================
Dep. Variable:         popularity_abs   R-squared:                       0.581
Model:                            OLS   Adj. R-squared:                  0.560
Method:                 Least Squares   F-statistic:                     28.04
Date:                Sun, 06 Jun 2021   Prob (F-statistic):           1.27e-14
Time:                        00:28:33   Log-Likelihood:                -278.85
No. Observations:                  86   AIC:                             567.7
Df Residuals:                      81   BIC:                             580.0
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
```

```
                  coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept       35.0641      3.231     10.851      0.000      28.635      41.494
track_number    -0.9542      0.201     -4.756      0.000      -1.353      -0.555
duration        21.9196      4.186      5.237      0.000      13.591      30.248
danceability    21.4110      5.942      3.603      0.001       9.589      33.233
age_days        -0.0047      0.001     -7.809      0.000      -0.006      -0.003
==============================================================================
Omnibus:                        2.863   Durbin-Watson:                   1.581
Prob(Omnibus):                  0.239   Jarque-Bera (JB):                2.782
Skew:                           0.384   Prob(JB):                        0.249
Kurtosis:                       2.569   Cond. No.                     2.49e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly spe
cified.
[2] The condition number is large, 2.49e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
--- SUMMARY FOR popularity_norm ~ track_number + duration + danceability + age_days
---
                            OLS Regression Results
==============================================================================
Dep. Variable:        popularity_norm   R-squared:                       0.581
Model:                            OLS   Adj. R-squared:                  0.560
Method:                 Least Squares   F-statistic:                     28.04
Date:                Sun, 06 Jun 2021   Prob (F-statistic):           1.27e-14
Time:                        00:28:33   Log-Likelihood:                 40.512
No. Observations:                  86   AIC:                            -71.02
Df Residuals:                      81   BIC:                            -58.75
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept        0.3430      0.079      4.352      0.000       0.186       0.500
track_number    -0.0233      0.005     -4.756      0.000      -0.033      -0.014
duration         0.5346      0.102      5.237      0.000       0.331       0.738
danceability     0.5222      0.145      3.603      0.001       0.234       0.811
age_days        -0.0001   1.46e-05     -7.809      0.000      -0.000   -8.47e-05
==============================================================================
Omnibus:                        2.863   Durbin-Watson:                   1.581
Prob(Omnibus):                  0.239   Jarque-Bera (JB):                2.782
Skew:                           0.384   Prob(JB):                        0.249
Kurtosis:                       2.569   Cond. No.                     2.49e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly spe
cified.
[2] The condition number is large, 2.49e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
--- SUMMARY FOR popularity_abs ~ track_number + duration + speechiness + age_days +
duration*complexity + danceability*valence + danceability*complexity ---
                            OLS Regression Results
==============================================================================
Dep. Variable:         popularity_abs   R-squared:                       0.664
Model:                            OLS   Adj. R-squared:                  0.620
Method:                 Least Squares   F-statistic:                     14.85
Date:                Sun, 06 Jun 2021   Prob (F-statistic):           3.58e-14
Time:                        00:28:33   Log-Likelihood:                -269.26
No. Observations:                  86   AIC:                             560.5
Df Residuals:                      75   BIC:                             587.5
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
=======
                             coef    std err          t      P>|t|      [0.025
0.975]
```

```
------------------------------------------------------------------------------
-------
Intercept                  42.0800      6.793      6.194      0.000      28.547
55.613
track_number               -0.9374      0.196     -4.789      0.000      -1.327
-0.547
duration                   58.5974     15.450      3.793      0.000      27.820
89.374
speechiness                11.5817      5.567      2.080      0.041       0.491
22.672
age_days                   -0.0048      0.001     -8.392      0.000      -0.006
-0.004
complexity                -37.0731     20.492     -1.809      0.074     -77.896
3.750
duration:complexity       -53.8892     17.620     -3.058      0.003     -88.991
-18.787
danceability              -43.2065     26.172     -1.651      0.103     -95.344
8.931
valence                    37.9028     19.364      1.957      0.054      -0.672
76.478
danceability:valence     -100.9028     45.160     -2.234      0.028    -190.866
-10.940
danceability:complexity   159.4315     48.251      3.304      0.001      63.311
255.552
==============================================================================
Omnibus:                        1.542   Durbin-Watson:                   1.642
Prob(Omnibus):                  0.463   Jarque-Bera (JB):                1.427
Skew:                           0.191   Prob(JB):                        0.490
Kurtosis:                       2.498   Cond. No.                     2.67e+05
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly spe
cified.
[2] The condition number is large, 2.67e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
--- SUMMARY FOR popularity_norm ~ track_number + duration + speechiness + age_days +
duration*complexity + danceability*valence + danceability*complexity ---
                            OLS Regression Results
==============================================================================
Dep. Variable:        popularity_norm   R-squared:                       0.664
Model:                            OLS   Adj. R-squared:                  0.620
Method:                 Least Squares   F-statistic:                     14.85
Date:                Sun, 06 Jun 2021   Prob (F-statistic):           3.58e-14
Time:                        00:28:33   Log-Likelihood:                 50.105
No. Observations:                  86   AIC:                            -78.21
Df Residuals:                      75   BIC:                            -51.21
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
=======
                             coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
-------
Intercept                  0.5141      0.166      3.103      0.003       0.184
0.844
track_number              -0.0229      0.005     -4.789      0.000      -0.032
-0.013
duration                   1.4292      0.377      3.793      0.000       0.679
2.180
speechiness                0.2825      0.136      2.080      0.041       0.012
0.553
age_days                  -0.0001    1.4e-05     -8.392      0.000      -0.000   -
8.95e-05
complexity                -0.9042      0.500     -1.809      0.074      -1.900
0.091
duration:complexity       -1.3144      0.430     -3.058      0.003      -2.171
-0.458
```

```
danceability               -1.0538      0.638     -1.651      0.103      -2.325
0.218
valence                     0.9245      0.472      1.957      0.054      -0.016
1.865
danceability:valence       -2.4610      1.101     -2.234      0.028      -4.655
-0.267
danceability:complexity     3.8886      1.177      3.304      0.001       1.544
6.233
==============================================================================
Omnibus:                        1.542   Durbin-Watson:                   1.642
Prob(Omnibus):                  0.463   Jarque-Bera (JB):                1.427
Skew:                           0.191   Prob(JB):                        0.490
Kurtosis:                       2.498   Cond. No.                     2.67e+05
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly spe
cified.
[2] The condition number is large, 2.67e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```