# Appendix 11: Hypothesis 3 Test

In [29]:

```
import pandas as pd
import numpy as np
import seaborn as sns
from sklearn.preprocessing import MinMaxScaler
%matplotlib inline
import matplotlib
import matplotlib.pyplot as plt
import statsmodels.api as sm
from scipy import stats

from statsmodels.formula.api import ols
```

In [30]:

```
df=pd.read_csv('data_final.csv')
df.head()
```

Out[30]:

| | Unnamed: 0 | track | artist | uri | danceability | energy |
|---|---|---|---|---|---|---|
| 0 | 0 | Wild Things | Alessia Cara | spotify:track:2ZyuwVvV6Z3XJaXIFbspeE | 0.741 | 0.62 |
| 1 | 1 | Love Someone | Lukas Graham | spotify:track:2JqnpexIO9dmvjUMCaLCLJ | 0.550 | 0.41 |
| 2 | 2 | Here's To Never Growing Up | Avril Lavigne | spotify:track:0qwcGscxUHGZTgq0zcaqk1 | 0.482 | 0.87 |
| 3 | 3 | Crawling Back To You | Daughtry | spotify:track:6BDtTzjbJ5kKKSWcJT8MIX | 0.438 | 0.91 |
| 4 | 4 | Faster | Matt Nathanson | spotify:track:6pIKFdrBnKF0y3CRuceTDh | 0.742 | 0.85 |

5 rows × 32 columns

In [31]:

```
df=df.iloc[:,1:]
df.head()
```

Out[31]:

| | track | artist | uri | danceability | energy | key | louc |
|---|---|---|---|---|---|---|---|
| 0 | Wild Things | Alessia Cara | spotify:track:2ZyuwVvV6Z3XJaXIFbspeE | 0.741 | 0.626 | 1 | - |
| 1 | Love Someone | Lukas Graham | spotify:track:2JqnpexIO9dmvjUMCaLCLJ | 0.550 | 0.415 | 9 | - |
| 2 | Here's To Never Growing Up | Avril Lavigne | spotify:track:0qwcGscxUHGZTgq0zcaqk1 | 0.482 | 0.873 | 0 | - |
| 3 | Crawling Back To You | Daughtry | spotify:track:6BDtTzjbJ5kKKSWcJT8MIX | 0.438 | 0.919 | 0 | - |
| 4 | Faster | Matt Nathanson | spotify:track:6pIKFdrBnKF0y3CRuceTDh | 0.742 | 0.853 | 9 | - |

5 rows × 31 columns

In [32]:

```
genres=df['genres'].unique()
decades=df['Decade'].unique()
print(genres)
print(decades)
```

```
['Pop' 'Easy listening' 'Hip hop' 'Metal' 'Country' 'Electronic' 'Rock'
 'Folk' 'Latin' 'Classical' 'Jazz' 'R&B' 'Caribbean' 'Blues']
['10s' '00s' '90s' '80s' '70s' '60s']
```

In [33]:

```
tb = [[0 for x in range(6)] for y in range(len(df))]
len(tb)
```

Out[33]:

```
33354
```

In [34]:

```
df_1=df
y=df_1.iloc[:,[25,26]]
x=df_1.iloc[:,[3,4,10,12]]
norm = MinMaxScaler().fit(x)
x_norm = norm.transform(x)
x=pd.DataFrame(data=x_norm,columns=x.columns)
x.head()
data=pd.concat([x,y],axis=1)
data.head()
```

Out[34]:

| | danceability | energy | instrumentalness | valence | genres | Decade |
|---|---|---|---|---|---|---|
| 0 | 0.734180 | 0.625906 | 0.000000 | 0.710977 | Pop | 10s |
| 1 | 0.528627 | 0.414853 | 0.000000 | 0.275932 | Pop | 10s |
| 2 | 0.455446 | 0.872968 | 0.000000 | 0.742195 | Pop | 10s |
| 3 | 0.408093 | 0.918980 | 0.000000 | 0.196375 | Pop | 10s |
| 4 | 0.735256 | 0.852963 | 0.000005 | 0.956697 | Pop | 10s |

In [35]:

```
for index,row in data.iterrows():
    tb[index][0]=row['genres']
    tb[index][1]=row['Decade']
    tb[index][2]=row['danceability']
    tb[index][3]=row['energy']
    tb[index][4]=row['instrumentalness']
    tb[index][5]=row['valence']
```

In [36]:

```
tb[:5]
```

Out[36]:

```
[['Pop',
  '10s',
  0.7341799397331037,
  0.6259061024317104,
  0.0,
  0.7109768378650553],
 ['Pop',
  '10s',
  0.5286267757210504,
  0.41485312813516195,
  0.0,
  0.27593152064451154],
 ['Pop',
  '10s',
  0.4554455445544555,
  0.8729681149968643,
  0.0,
  0.7421953675730111],
 ['Pop',
  '10s',
  0.40809298321136456,
  0.9189796638956379,
  0.0,
  0.19637462235649547],
 ['Pop',
  '10s',
  0.735256134309083,
  0.8529630937365279,
  4.79e-06,
  0.9566968781470291]]
```

In [38]:

```
x=['Genre','Decade','Danceability','Energy','Instrumentalness','Valence']
df_tb=pd.DataFrame(data=tb,columns=x)
df_tb.head()
```

Out[38]:

| | Genre | Decade | Danceability | Energy | Instrumentalness | Valence |
|---|---|---|---|---|---|---|
| 0 | Pop | 10s | 0.734180 | 0.625906 | 0.000000 | 0.710977 |
| 1 | Pop | 10s | 0.528627 | 0.414853 | 0.000000 | 0.275932 |
| 2 | Pop | 10s | 0.455446 | 0.872968 | 0.000000 | 0.742195 |
| 3 | Pop | 10s | 0.408093 | 0.918980 | 0.000000 | 0.196375 |
| 4 | Pop | 10s | 0.735256 | 0.852963 | 0.000005 | 0.956697 |

In [39]:

```
d_melt=df_tb.iloc[:,:3]
d_melt.head()
```

Out[39]:

| | Genre | Decade | Danceability |
|---|---|---|---|
| 0 | Pop | 10s | 0.734180 |
| 1 | Pop | 10s | 0.528627 |
| 2 | Pop | 10s | 0.455446 |
| 3 | Pop | 10s | 0.408093 |
| 4 | Pop | 10s | 0.735256 |

In [41]:

```
model = ols('Danceability ~ C(Genre) + C(Decade) + C(Genre):C(Decade)', data=d_melt).fi
t()
anova_table = sm.stats.anova_lm(model, typ=2)
anova_table
```

Out[41]:

| | sum_sq | df | F | PR(>F) |
|---|---|---|---|---|
| C(Genre) | 335.326773 | 13.0 | 1080.019062 | 0.000000e+00 |
| C(Decade) | 27.072615 | 5.0 | 226.707950 | 8.732465e-239 |
| C(Genre):C(Decade) | 32.805159 | 65.0 | 21.131743 | 7.966763e-239 |
| Residual | 794.595785 | 33270.0 | NaN | NaN |

In [43]:

```
d_melt=df_tb.iloc[:,[0,1,3]]
model = ols('Energy ~ C(Genre) + C(Decade) + C(Genre):C(Decade)', data=d_melt).fit()
anova_table = sm.stats.anova_lm(model, typ=2)
anova_table
```

Out[43]:

| | sum_sq | df | F | PR(>F) |
|---|---|---|---|---|
| C(Genre) | 488.358079 | 13.0 | 999.265824 | 0.0 |
| C(Decade) | 101.612683 | 5.0 | 540.584918 | 0.0 |
| C(Genre):C(Decade) | 72.445538 | 65.0 | 29.647242 | 0.0 |
| Residual | 1250.739285 | 33270.0 | NaN | NaN |

In [44]:

```
d_melt=df_tb.iloc[:,[0,1,4]]
model = ols('Instrumentalness ~ C(Genre) + C(Decade) + C(Genre):C(Decade)', data=d_melt
).fit()
anova_table = sm.stats.anova_lm(model, typ=2)
anova_table
```

Out[44]:

| | sum_sq | df | F | PR(>F) |
|---|---|---|---|---|
| C(Genre) | 815.699946 | 13.0 | 1054.430106 | 0.000000e+00 |
| C(Decade) | 15.742111 | 5.0 | 52.908285 | 6.935869e-55 |
| C(Genre):C(Decade) | 98.902500 | 65.0 | 25.569641 | 3.059811e-296 |
| Residual | 1979.803487 | 33270.0 | NaN | NaN |

In [45]:

```
d_melt=df_tb.iloc[:,[0,1,5]]
model = ols('Valence ~ C(Genre) + C(Decade) + C(Genre):C(Decade)', data=d_melt).fit()
anova_table = sm.stats.anova_lm(model, typ=2)
anova_table
```

Out[45]:

| | sum_sq | df | F | PR(>F) |
|---|---|---|---|---|
| C(Genre) | 417.757155 | 13.0 | 587.067008 | 0.000000e+00 |
| C(Decade) | 51.910606 | 5.0 | 189.667631 | 6.678632e-200 |
| C(Genre):C(Decade) | 42.290686 | 65.0 | 11.886076 | 5.125215e-120 |
| Residual | 1821.149802 | 33270.0 | NaN | NaN |

In [ ]:

```

```