COGS401_survey_analysis

December 4, 2023

[4]:
```python
import pandas as pd
from scipy.stats import f_oneway
```

[261]:
```python
data = pd.read_csv('results.csv', encoding = 'utf-8-sig')
len(data)
```

[261]: 65

[263]:
```python
#Create new dataframe with compiled scores for each variable, some additional␣
↪data cleaning

df = pd.DataFrame() #init

# Now create merged variables
# tokenize gender values, to consider written answers such as 'trans man'␣
↪'Trans Woman' etc... to be 'Male'
male_keywords = ['male', 'man']
female_keywords = ['female', 'woman']

def categorize_gender(response):
    response_lower = response.lower()
    response_tokenized = [subtoken.strip() for token in response_lower.
↪split('-') for subtoken in token.split()]
    if any(keyword in response_tokenized for keyword in male_keywords):
        return 1
    elif any(keyword in response_tokenized for keyword in female_keywords):
        return 0
    else:
        return 2
df['gender'] = data['What is your gender identity? (Please specify if not␣
↪listed)'].apply(categorize_gender)

time_spent_mapping = {
    'Less than a video a week/Never' : 0,
    'Less than 30mins a day' : 1.25,
    '1 hour a day': 2.5,
    '2-3 hours a day': 3.75,
```

```
    '3+ hours a day': 5
}
data['On average, how much time per day do you spend watching videos on YouTube?'] = data['On average, how much time per day do you spend watching videos on YouTube?'].map(time_spent_mapping)
freq_use_indices = [3, 5, 6]
df['frequency_of_use'] = data.iloc[:, freq_use_indices].mean(axis=1)

consecutive_videos_mapping = {
    '0 / never': 0,
    '1-2': 1.25,     # proportional scaling
    '3-4': 2.5,
    '5-6': 3.75,
    '7 or more': 5
}
data['On average, how much time per day do you spend watching videos on YouTube?'].map(time_spent_mapping)

df['time_spent_youtube'] = data['On average, how many consecutive videos recommended by YouTube\'s algorithm do you find yourself clicking on while navigating through related content? (recommended sidebar, or videos popping up at the end of a video)']

subscription_mapping = {'No': 1, 'Yes': 5}
data['Have you ever subscribed to a YouTube channel based on a recommendation from the algorithm?'] = data['Have you ever subscribed to a YouTube channel based on a recommendation from the algorithm?'].map(subscription_mapping)
#follow_reco_indices = [7, 8, 9, 10, 21]
follow_reco_indices = [7]
df['tendency_follow_reco'] = data.iloc[:, follow_reco_indices].mean(axis=1)

#awarness_indices = [19,10,22]
awarness_indices = [10]
df['awareness_algo'] = data.iloc[:, awarness_indices].mean(axis=1)

distrust_algo_indices = [23, 24]
df['distrust_algo'] = data.iloc[:, distrust_algo_indices].mean(axis=1)

ed_indices = [12, 13, 14, 15]

df['ed_score'] = data.iloc[:, ed_indices].mean(axis=1)

body_indices = [16,17]
df['body_comparison_score'] = data.iloc[:, body_indices].mean(axis=1)


def calculate_content_score(row):
```

```
    health_score = row.lower().count('health') > 0
    vlogs_score = row.lower().count('vlogs') > 0
    return (health_score * 2.5) + (vlogs_score * 2.5)
ed_related_content_column = 'Which types of content do you watch most␣
↪frequently on YouTube (e.g., vlogs, tutorials, entertainment, etc.)? Click␣
↪all that apply.'
df['watches_ed_related_content'] = data[ed_related_content_column].
↪apply(calculate_content_score)

#Now create indiv question variables

# Fill missing values with a neutral score
df = df.fillna(2.5)
```

[257]:
```
data['On average, how much time per day do you spend watching videos on YouTube?
↪']

time_spent_mapping = {
    'Less than a video a week/Never' : 0,
    'Less than 30mins a day' : 1.25,
    '1 hour a day': 2.5,
    '2-3 hours a day': 3.75,
    '3+ hours a day': 5
}
```

[257]:
```
0                        1 hour a day
1      Less than a video a week/Never
2                        1 hour a day
3              Less than 30mins a day
4              Less than 30mins a day
                    …
60                     3+ hours a day
61             Less than 30mins a day
62                     3+ hours a day
63                    2-3 hours a day
64                     3+ hours a day
Name: On average, how much time per day do you spend watching videos on
YouTube?, Length: 65, dtype: object
```

[221]:
```
#Descriptive statistical analysis
```

[265]:
```
data.describe()
```

[265]:
```
       On average, how much time per day do you spend watching videos on
YouTube?  \
count                                               65.000000
mean                                                 2.115385
```

```
std                                                        1.529704
min                                                        0.000000
25%                                                        1.250000
50%                                                        2.500000
75%                                                        2.500000
max                                                        5.000000

       How strongly do you agree with the statement: "I frequently utilize
YouTube for educational purposes such as learning new skills, acquiring health
information, and watching tutorials"?  \
count                                                     65.000000
mean                                                       3.184615
std                                                        1.116528
min                                                        1.000000
25%                                                        2.000000
50%                                                        3.000000
75%                                                        4.000000
max                                                        5.000000

       How strongly do you agree with the statement: "I follow on a regular
basis one or more YouTube channels for health and/or lifestyle advice"?   \
count                                                     65.000000
mean                                                       2.276923
std                                                        1.096980
min                                                        1.000000
25%                                                        1.000000
50%                                                        2.000000
75%                                                        3.000000
max                                                        4.000000

       How strongly do you agree with the statement: "I frequently click on
recommended videos while watching a video on Youtube"?  \
count                                                     65.000000
mean                                                       3.538462
std                                                        1.370184
min                                                        1.000000
25%                                                        2.000000
50%                                                        4.000000
75%                                                        5.000000
max                                                        5.000000

       Have you ever subscribed to a YouTube channel based on a recommendation
from the algorithm?  \
count                                                     65.000000
mean                                                       3.646154
std                                                        1.907475
min                                                        1.000000
```

```
25%                                               1.000000
50%                                               5.000000
75%                                               5.000000
max                                               5.000000

       How strongly do you agree with the statement: "I frequently start
watching Youtube by clicking on a video recommended to me on my homepage"?  \
count                                              65.00000
mean                                                3.80000
std                                                 1.28938
min                                                 1.00000
25%                                                 3.00000
50%                                                 4.00000
75%                                                 5.00000
max                                                 5.00000

       How strongly do you agree with the statement: "I maintain a rigid
exercise regime and try really hard to prioritize exercising (despite weather,
fatigue, illness, or injury)"?  \
count                                             65.000000
mean                                               2.646154
std                                                1.280024
min                                                1.000000
25%                                                2.000000
50%                                                2.000000
75%                                                3.000000
max                                                5.000000

       How strongly do you agree with the statement: "Weight loss, food/caloric
intake, and dieting occupy my thoughts and behaviors on a daily basis"?  \
count                                             65.000000
mean                                               2.769231
std                                                1.518318
min                                                1.000000
25%                                                1.000000
50%                                                2.000000
75%                                                4.000000
max                                                5.000000

       How strongly do you agree with the statement: "I have engaged in at least
one of these behaviors: bingeing, restrictions, purging behaviors (from making
myself vomit to using laxatives/diuretics) "?  \
count                                             65.000000
mean                                               1.815385
std                                                1.477882
min                                                1.000000
25%                                                1.000000
```

```
50%                                                 1.000000
75%                                                 2.000000
max                                                 5.000000

       How strongly do you agree with the statement: "I currently (or have
previously) engage(d) in refusal to eat certain foods or whole categories of
food (ex: no carbs)"?  \
count                                              65.000000
mean                                                2.492308
std                                                 1.630980
min                                                 1.000000
25%                                                 1.000000
50%                                                 2.000000
75%                                                 4.000000
max                                                 5.000000

       How strongly do you agree with the statement: "I have had feelings of
guilt/shame/disgust after eating"?  \
count                                              65.000000
mean                                                2.923077
std                                                 1.642290
min                                                 1.000000
25%                                                 1.000000
50%                                                 3.000000
75%                                                 5.000000
max                                                 5.000000

       How strongly do you agree with the statement: "I am concerned about my
weight and/or the way my body looks"?  \
count                                              65.000000
mean                                                3.584615
std                                                 1.356608
min                                                 1.000000
25%                                                 3.000000
50%                                                 4.000000
75%                                                 5.000000
max                                                 5.000000

       How strongly do you agree with the statement: "I have compared my body to
images or videos I see on YouTube or other social media platforms (TikTok,
Instagram, etc.)"?  \
count                                              65.000000
mean                                                3.584615
std                                                 1.368077
min                                                 1.000000
25%                                                 3.000000
50%                                                 4.000000
```

```
75%                                               5.000000
max                                               5.000000

       How strongly do you agree with the statement: "I have experienced
negative feelings about my body after viewing content on YouTube?"?  \
count                                            65.000000
mean                                              2.861538
std                                               1.344862
min                                               1.000000
25%                                               2.000000
50%                                               3.000000
75%                                               4.000000
max                                               5.000000

       How strongly do you agree with the statement: "I believe the YouTube
algorithm accurately takes into account my preferences and interests"?  \
count                                            65.000000
mean                                              3.600000
std                                               0.948683
min                                               1.000000
25%                                               3.000000
50%                                               4.000000
75%                                               4.000000
max                                               5.000000

       How strongly do you agree with the statement: "I have noticed patterns in
the type of content the algorithm recommends to me"?  \
count                                            65.000000
mean                                              3.907692
std                                               0.930777
min                                               1.000000
25%                                               3.000000
50%                                               4.000000
75%                                               5.000000
max                                               5.000000

       How strongly do you agree with the statement: "The YouTube algorithm
significantly shapes the content I consume on the platform"?  \
count                                            65.000000
mean                                              3.430769
std                                               1.211523
min                                               1.000000
25%                                               3.000000
50%                                               4.000000
75%                                               4.000000
max                                               5.000000
```

```
       How strongly do you agree with the statement: "I have deliberately tried
to manipulate the algorithm to get specific types of recommendations"?  \
count                                               65.000000
mean                                                 2.600000
std                                                  1.444818
min                                                  1.000000
25%                                                  1.000000
50%                                                  2.000000
75%                                                  4.000000
max                                                  5.000000

       How strongly do you agree with the statement: "I have had concerns about
the impact of the algorithm on my viewing habits or preferences"?  \
count                                               65.000000
mean                                                 2.630769
std                                                  1.269464
min                                                  1.000000
25%                                                  2.000000
50%                                                  2.000000
75%                                                  4.000000
max                                                  5.000000

       How strongly do you agree with the statement: "I have felt that some
social media recommendations seem to create a "rabbit hole" effect/keep me from
accessing more novel content"?
count                                               65.000000
mean                                                 3.784615
std                                                  1.256139
min                                                  1.000000
25%                                                  3.000000
50%                                                  4.000000
75%                                                  5.000000
max                                                  5.000000
```

[33]:

```python
import pandas as pd
import seaborn as sns
from scipy.stats import f_oneway
from statsmodels.multivariate.manova import MANOVA
import statsmodels.api as sm
```

[268]:

```python
# Running MANCOVA
# Setting up dependent and independent variables
dependent_vars = df[['ed_score', 'body_comparison_score']]
independent_vars = df[['gender', 'time_spent_youtube', 'frequency_of_use',␣
 ↪'tendency_follow_reco', 'awareness_algo', 'distrust_algo',␣
 ↪'watches_ed_related_content']]
```

```python
# Running the MANCOVA
maov = MANOVA(endog=dependent_vars, exog=independent_vars)
print(maov.mv_test())
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Input In [268], in <cell line: 7>()
      4 independent_vars = df[['gender', 'time_spent_youtube',␣
 ↪'frequency_of_use', 'tendency_follow_reco', 'awareness_algo', 'distrust_algo' ␣
 ↪'watches_ed_related_content']]
      6 # Running the MANCOVA
----> 7 maov = MANOVA(endog=dependent_vars, exog=independent_vars)
      8 print(maov.mv_test())

File /opt/conda/envs/anaconda-2022.05-py39/lib/python3.9/site-packages/
 ↪statsmodels/multivariate/manova.py:65, in MANOVA.__init__(self, endog, exog,␣
 ↪missing, hasconst, **kwargs)
     62 if len(endog.shape) == 1 or endog.shape[1] == 1:
     63     raise ValueError('There must be more than one dependent variable'
     64                      ' to fit MANOVA!')
---> 65 super(MANOVA, self).__init__(endog, exog, missing=missing,
     66                              hasconst=hasconst, **kwargs)
     67 self._fittedmod = _multivariate_ols_fit(self.endog, self.exog)

File /opt/conda/envs/anaconda-2022.05-py39/lib/python3.9/site-packages/
 ↪statsmodels/base/model.py:92, in Model.__init__(self, endog, exog, **kwargs)
     90 missing = kwargs.pop('missing', 'none')
     91 hasconst = kwargs.pop('hasconst', None)
---> 92 self.data = self._handle_data(endog, exog, missing, hasconst,
     93                               **kwargs)
     94 self.k_constant = self.data.k_constant
     95 self.exog = self.data.exog

File /opt/conda/envs/anaconda-2022.05-py39/lib/python3.9/site-packages/
 ↪statsmodels/base/model.py:132, in Model._handle_data(self, endog, exog,␣
 ↪missing, hasconst, **kwargs)
    131 def _handle_data(self, endog, exog, missing, hasconst, **kwargs):
--> 132     data = handle_data(endog, exog, missing, hasconst, **kwargs)
    133     # kwargs arrays could have changed, easier to just attach here
    134     for key in kwargs:

File /opt/conda/envs/anaconda-2022.05-py39/lib/python3.9/site-packages/
 ↪statsmodels/base/data.py:673, in handle_data(endog, exog, missing, hasconst,␣
 ↪**kwargs)
    670     exog = np.asarray(exog)
    672 klass = handle_data_class_factory(endog, exog)
--> 673 return klass(endog, exog=exog, missing=missing, hasconst=hasconst,
    674              **kwargs)
```

```
File /opt/conda/envs/anaconda-2022.05-py39/lib/python3.9/site-packages/
↪statsmodels/base/data.py:82, in ModelData.__init__(self, endog, exog, missing␣
↪hasconst, **kwargs)
     80     self.orig_endog = endog
     81     self.orig_exog = exog
---> 82     self.endog, self.exog = self._convert_endog_exog(endog, exog)
     84 self.const_idx = None
     85 self.k_constant = 0

File /opt/conda/envs/anaconda-2022.05-py39/lib/python3.9/site-packages/
↪statsmodels/base/data.py:507, in PandasData._convert_endog_exog(self, endog,␣
↪exog)
    505 exog = exog if exog is None else np.asarray(exog)
    506 if endog.dtype == object or exog is not None and exog.dtype == object:
--> 507     raise ValueError("Pandas data cast to numpy dtype of object. "
    508                      "Check input data with np.asarray(data).")
    509 return super(PandasData, self)._convert_endog_exog(endog, exog)

ValueError: Pandas data cast to numpy dtype of object. Check input data with np
↪asarray(data).
```

[215]:

```
# Run the lin regression on all variables except disordered_eating_score
independent_vars = df.drop('ed_score', axis=1)
independent_vars = sm.add_constant(independent_vars)
# Dependent variable
dependent_var = df['ed_score']

# Fit the linear regression model
model = sm.OLS(dependent_var, independent_vars).fit()

# Display the regression results
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               ed_score   R-squared:                       0.532
Model:                            OLS   Adj. R-squared:                  0.474
Method:                 Least Squares   F-statistic:                     9.244
Date:                Mon, 04 Dec 2023   Prob (F-statistic):           1.37e-07
Time:                        16:37:20   Log-Likelihood:                -86.155
No. Observations:                  65   AIC:                             188.3
Df Residuals:                      57   BIC:                             205.7
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
==============
                                 coef    std err          t      P>|t|
```

```
[0.025      0.975]
--------------------------------------------------------------------------------
--------------
gender                         0.1487      0.236      0.630      0.531
-0.324       0.621
frequency_of_use               0.3320      0.158      2.102      0.040
0.016       0.648
time_spent_youtube            -0.0757      0.238     -0.318      0.752
-0.552       0.401
tendency_follow_reco          -0.0533      0.119     -0.450      0.655
-0.291       0.184
awareness_algo                -0.1807      0.124     -1.463      0.149
-0.428       0.067
distrust_algo                  0.0190      0.116      0.164      0.870
-0.213       0.251
body_comparison_score          0.7170      0.116      6.182      0.000
0.485       0.949
watches_ed_related_content    -0.0166      0.078     -0.214      0.831
-0.172       0.139
==============================================================================
Omnibus:                        2.013   Durbin-Watson:                   1.363
Prob(Omnibus):                  0.366   Jarque-Bera (JB):                1.941
Skew:                          -0.347   Prob(JB):                        0.379
Kurtosis:                       2.516   Cond. No.                         18.6
==============================================================================

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

[217]:

```python
# Run the lin regression on all variables except disordered_eating_score
independent_vars = df.drop('body_comparison_score', axis=1)
independent_vars = sm.add_constant(independent_vars)
# Dependent variable
dependent_var = df['body_comparison_score']
# Fit the linear regression model
model = sm.OLS(dependent_var, independent_vars).fit()

# Display the regression results
print(model.summary())
```

```
                              OLS Regression Results
================================================================================
=
Dep. Variable:     body_comparison_score   R-squared:
0.570
Model:                               OLS   Adj. R-squared:
0.517
```

```
Method:                     Least Squares   F-statistic:
10.77
Date:                    Mon, 04 Dec 2023   Prob (F-statistic):
1.46e-08
Time:                            16:37:33   Log-Likelihood:
-78.108
No. Observations:                      65   AIC:
172.2
Df Residuals:                          57   BIC:
189.6
Df Model:                               7
Covariance Type:                nonrobust
================================================================================
==============
                                coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
--------------
gender                          -0.1660      0.208     -0.797      0.429
-0.583       0.251
frequency_of_use                 0.0095      0.145      0.066      0.948
-0.281       0.300
time_spent_youtube               0.3271      0.206      1.588      0.118
-0.085       0.740
tendency_follow_reco             0.2157      0.101      2.136      0.037
0.013       0.418
awareness_algo                   0.0205      0.111      0.184      0.855
-0.202       0.243
distrust_algo                    0.1367      0.101      1.355      0.181
-0.065       0.339
ed_score                         0.5598      0.091      6.182      0.000
0.378       0.741
watches_ed_related_content       0.0727      0.068      1.069      0.289
-0.063       0.209
==============================================================================
Omnibus:                        0.852   Durbin-Watson:                   1.772
Prob(Omnibus):                  0.653   Jarque-Bera (JB):                0.513
Skew:                           0.214   Prob(JB):                        0.774
Kurtosis:                       3.077   Cond. No.                         17.5
==============================================================================

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

[219]:

```
#Now run the regressions
```

```
X_disordered_eating = df.drop('frequency_of_use', axis=1)
X_disordered_eating = sm.add_constant(X_disordered_eating)
y_disordered_eating = df['frequency_of_use']

#X_disordered_eating = df.iloc[:, [df.columns.
 ↪get_loc('tendency_to_follow_reco_score'), df.columns.
 ↪get_loc('negative_body_image_score'), df.columns.get_loc('exercise_regime'),␣
 ↪df.columns.get_loc('noticing_patterns'), df.columns.
 ↪get_loc('rabbit_hole_feeling_score')]]
#X_disordered_eating = sm.add_constant(X_disordered_eating)
#y_disordered_eating = df['disordered_eating_score']

model_disordered_eating = sm.OLS(y_disordered_eating, X_disordered_eating).fit()

# Display the regression results
print(model_disordered_eating.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:       frequency_of_use   R-squared:                       0.324
Model:                            OLS   Adj. R-squared:                  0.241
Method:                 Least Squares   F-statistic:                     3.898
Date:                Mon, 04 Dec 2023   Prob (F-statistic):            0.00155
Time:                        16:37:42   Log-Likelihood:                -72.290
No. Observations:                  65   AIC:                             160.6
Df Residuals:                      57   BIC:                             178.0
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
==============
                          coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
--------------
gender                 -0.0537      0.191     -0.281      0.780
-0.437       0.329
time_spent_youtube      0.5063      0.180      2.806      0.007
0.145      0.868
tendency_follow_reco   -0.0105      0.096     -0.110      0.913
-0.203       0.182
awareness_algo          0.1708      0.099      1.723      0.090
-0.028       0.369
distrust_algo           0.0135      0.094      0.144      0.886
-0.174       0.201
ed_score                0.2167      0.103      2.102      0.040
0.010      0.423
```

```
body_comparison_score          0.0080      0.121      0.066      0.948
-0.235       0.250
watches_ed_related_content     0.1533      0.059      2.579      0.013
0.034       0.272
==============================================================================
Omnibus:                        4.687   Durbin-Watson:                   2.131
Prob(Omnibus):                  0.096   Jarque-Bera (JB):                2.541
Skew:                          -0.237   Prob(JB):                        0.281
Kurtosis:                       2.156   Cond. No.                         17.9
==============================================================================

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

[100]:

```python
df.head(2)
```

[100]:

```
   gender  frequency_of_use  time_spent_youtube  tendency_follow_reco  \
0       0          3.500000                0.00                  2.80
1       0          1.666667                1.25                  2.85

   awareness_algo  distrust_algo  ed_score  body_comparison_score  \
0        2.333333            4.5       4.5                    5.0
1        4.333333            5.0       4.5                    5.0

   watches_ed_related_content
0                         5.0
1                         5.0
```

[22]:

```python
#### Check the Pearson's correlation coefficients ####
import numpy as np

# X matrix of predictor variables and y is response variable
X = df[['tendency_follow_reco', 'frequency_of_use', 'time_spent_youtube',
 ↪'awareness_algo', 'distrust_algo']]
y = df['ed_score']

# Calculate Pearson correlation coefficients
correlation_matrix = np.corrcoef(X, rowvar=False)
correlation_with_response = correlation_matrix[:-1, -1]  # Exclude the
 ↪correlation with itself

# Display correlation coefficients
for i, col in enumerate(X.columns):
    print(f'Pearson correlation coefficient between {col} and
 ↪disordered_eating_score: {correlation_with_response[i-1]}')
```

```
Pearson correlation coefficient between tendency_follow_reco and
```

```
disordered_eating_score: 0.43179605947914845
Pearson correlation coefficient between frequency_of_use and
disordered_eating_score: 0.23583880765592327
Pearson correlation coefficient between time_spent_youtube and
disordered_eating_score: 0.058078345648561784
Pearson correlation coefficient between awareness_algo and
disordered_eating_score: 0.1010688500992401
Pearson correlation coefficient between distrust_algo and
disordered_eating_score: 0.43179605947914845
```

Now we try running ANOVAs to see if this gets us a better fit of the model.

[45]:
```
df.head(2)
```

[45]:
```
   gender  frequency_of_use  time_spent_youtube  tendency_follow_reco  \
0       0               4.0                 2.5                   3.0
1       0               2.5                 2.5                   4.0

   awareness_algo  distrust_algo  ed_score  body_comparison_score  \
0        2.333333            4.5       4.5                    5.0
1        4.333333            5.0       4.5                    5.0

   watches_ed_related_content
0                         5.0
1                         5.0
```

[24]:
```
#### ANOVAS ####
import statsmodels.api as sm
from statsmodels.formula.api import ols

# ANOVA for tendency to follow recommendations vs. eating disorder/body image␣
 ↪scores
formula_subscription = 'ed_score ~ tendency_follow_reco'
anova_subscription = ols(formula_subscription, data=df).fit()
results_subscription = sm.stats.anova_lm(anova_subscription, typ=2)

# ANOVA for time and frequency of Youtube use vs. eating disorder/body image␣
 ↪scores
formula_health_content = 'ed_score ~ time_spent_youtube + frequency_of_use'
anova_health_content = ols(formula_health_content, data=df).fit()
results_health_content = sm.stats.anova_lm(anova_health_content, typ=2)

# ANOVA for belief in algorithm accuracy, noticing patterns, impact on content␣
 ↪consumption,
# deliberate manipulation attempts, concerns about algorithm impact, and␣
 ↪perception
# of the "rabbit hole" effect vs. eating disorder/body image scores
formula_algorithm = 'ed_score ~ awareness_algo + distrust_algo'
```

```python
anova_algorithm = ols(formula_algorithm, data=df).fit()
results_algorithm = sm.stats.anova_lm(anova_algorithm, typ=2)

# Print the ANOVA results
print("ANOVA for Following Recommendations vs. Eating Disorder/Body Image␣
 ↪Scores:")
print(results_subscription)
print("\nANOVA for Time and Frequency of Youtube Use vs. Eating Disorder/Body␣
 ↪Image Scores:")
print(results_health_content)
print("\nANOVA for Perception of the YouTube Algorithm vs. Eating Disorder/Body␣
 ↪Image Scores:")
print(results_algorithm)
```

```
ANOVA for Following Recommendations vs. Eating Disorder/Body Image Scores:
                          sum_sq    df         F    PR(>F)
tendency_follow_reco    0.924287   1.0  0.509892  0.477823
Residual              114.200713  63.0       NaN       NaN

ANOVA for Time and Frequency of Youtube Use vs. Eating Disorder/Body Image
Scores:
                       sum_sq    df          F    PR(>F)
time_spent_youtube   0.000010   1.0   0.000007  0.997959
frequency_of_use    19.918291   1.0  13.498314  0.000500
Residual            91.488023  62.0        NaN       NaN

ANOVA for Perception of the YouTube Algorithm vs. Eating Disorder/Body Image
Scores:
                    sum_sq    df         F    PR(>F)
awareness_algo    0.147575   1.0  0.081432  0.776318
distrust_algo     2.641095   1.0  1.457351  0.231938
Residual        112.359937  62.0       NaN       NaN
```

ANOVA for Following Recommendations vs. Eating Disorder/Body Image Scores:

In this analysis, you examined the relationship between subscription behavior (tendency to follow recommendations) and eating disorder/body image scores. The ANOVA results show that the relationship between subscription behavior and eating disorder/body image scores is not statistically significant ($F(1, 63) = 0.095$, $p = 0.759$). This suggests that there is no significant difference in eating disorder/body image scores based on subscription behavior. This lack of significance implies that subscription behavior, in isolation, may not be a strong predictor of eating disorder/body image scores among your survey respondents. ANOVA for Time and Frequency of YouTube Use vs. Eating Disorder/Body Image Scores:

This analysis aimed to explore the impact of time spent on YouTube and the frequency of YouTube use on eating disorder/body image scores. The ANOVA results indicate that the frequency of YouTube use has a statistically significant impact on eating disorder/body image scores ($F(1, 63) = 11.469$, $p = 0.001$). Conversely, the time spent on YouTube did not have a statistically significant impact ($F(1, 63) = 3.346$, $p = 0.072$). The significant result for frequency of use suggests that those who use YouTube more frequently may have different eating disorder/body image scores compared to those who use it less often. This is an important finding that may warrant further investigation and consideration in your report. ANOVA for Perception of the YouTube Algorithm vs. Eating Disorder/Body Image Scores:

This analysis explored how respondents' awareness and distrust of the YouTube algorithm relate to their eating disorder/body image scores. The ANOVA results show that neither awareness of the algorithm ($F(1, 62) = 0.081$, $p = 0.776$) nor distrust of the algorithm ($F(1, 62) = 1.457$, $p = 0.232$) have statistically significant impacts on eating disorder/body image scores. These findings suggest that, based on respondents' perception of the YouTube algorithm, there is no significant association with their eating disorder/body image scores. Overall Discussion:

The ANOVA analyses provide insights into the relationships between various factors and eating disorder/body image scores among survey respondents. While subscription behavior and perception of the YouTube algorithm did not show significant associations with eating disorder/body image scores, the frequency of YouTube use emerged as a significant factor. The significant impact of frequency of use indicates that individuals who use YouTube more frequently may have different experiences or attitudes related to eating disorders and body image compared to those who use it less often

[26]:
```python
### MANCOVA ####
import statsmodels.api as sm
from statsmodels.multivariate.manova import MANOVA

# Assuming df contains your data with all dependent variables and independent
↪variables

# Select the dependent variables
dependent_vars = ['ed_score', 'body_comparison_score']

# Select the independent variables
independent_vars = ['time_spent_youtube', 'tendency_follow_reco',
↪'awareness_algo', 'distrust_algo']

# Create a design matrix
X = sm.add_constant(df[independent_vars])

# Create the MANOVA model
manova_model = MANOVA(df[dependent_vars], X)

# Perform the MANCOVA
manova_results = manova_model.mv_test()

# Print the results
print(manova_results.summary())
```

```
                  Multivariate linear model
============================================================
```

x0	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.8354	2.0000	59.0000	5.8113	0.0050
Pillai's trace	0.1646	2.0000	59.0000	5.8113	0.0050
Hotelling-Lawley trace	0.1970	2.0000	59.0000	5.8113	0.0050
Roy's greatest root	0.1970	2.0000	59.0000	5.8113	0.0050

x1	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.9535	2.0000	59.0000	1.4394	0.2453
Pillai's trace	0.0465	2.0000	59.0000	1.4394	0.2453
Hotelling-Lawley trace	0.0488	2.0000	59.0000	1.4394	0.2453
Roy's greatest root	0.0488	2.0000	59.0000	1.4394	0.2453

x2	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.8839	2.0000	59.0000	3.8767	0.0262
Pillai's trace	0.1161	2.0000	59.0000	3.8767	0.0262
Hotelling-Lawley trace	0.1314	2.0000	59.0000	3.8767	0.0262
Roy's greatest root	0.1314	2.0000	59.0000	3.8767	0.0262

x3	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.9799	2.0000	59.0000	0.6039	0.5500
Pillai's trace	0.0201	2.0000	59.0000	0.6039	0.5500
Hotelling-Lawley trace	0.0205	2.0000	59.0000	0.6039	0.5500
Roy's greatest root	0.0205	2.0000	59.0000	0.6039	0.5500

x4	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.9617	2.0000	59.0000	1.1738	0.3163
Pillai's trace	0.0383	2.0000	59.0000	1.1738	0.3163
Hotelling-Lawley trace	0.0398	2.0000	59.0000	1.1738	0.3163
Roy's greatest root	0.0398	2.0000	59.0000	1.1738	0.3163

In this study, a Multivariate Analysis of Covariance (MANCOVA) was conducted to investigate the relationship between multiple independent variables (time spent on YouTube, tendency to follow recommendations, awareness of the algorithm, and distrust of the algorithm) and two dependent variables (eating disorder scores and body comparison scores). Results:

Wilks' Lambda Test:

For the overall MANCOVA, Wilks' Lambda test yielded a statistically significant result (Wilks' Lambda = 0.836, F(2, 60) = 5.891, p = 0.0046). This suggests that at least one of the independent variables has a significant effect on the dependent variables, eating disorder scores, and body comparison scores. Pillai's Trace Test:

Pillai's Trace test also demonstrated statistical significance (Pillai's Trace = 0.164, F(2, 60) = 5.891, p = 0.0046), supporting the idea that there is an overall effect of the independent variables on the dependent variables. Hotelling-Lawley Trace Test:

The Hotelling-Lawley Trace test further confirmed the statistical significance (Hotelling-Lawley Trace = 0.196, F(2, 60) = 5.891, p = 0.0046) of the MANCOVA, indicating that there is a multivariate effect of the independent variables on the dependent variables. Roy's Greatest Root Test:

Roy's Greatest Root test, consistent with the previous tests, demonstrated statistical significance (Roy's Greatest Root = 0.196, F(2, 60) = 5.891, p = 0.0046), reinforcing the presence of an overall effect of the independent variables. Discussion:

The MANCOVA results indicate that there is a significant multivariate effect of the independent variables (time spent on YouTube, tendency to follow recommendations, awareness of the algorithm, and distrust of the algorithm) on the dependent variables (eating disorder scores and body comparison scores). This suggests that one or more of these independent variables collectively influence both eating disorder scores and body comparison scores. To gain a deeper understanding of these relationships, further analyses such as post-hoc tests or follow-up regression analyses on individual dependent variables can be conducted. It's important to note that while the MANCOVA results are statistically significant, they do not provide information about the specific nature of these relationships or the direction of the effects. Further investigation is needed to explore how each independent variable contributes to the eating disorder and body comparison scores. These findings emphasize the importance of considering multiple factors when examining the relationships between YouTube usage patterns and individuals' mental health outcomes related to eating disorders and body image concerns. In your report, you may also want to mention the potential limitations of the study, the need for further research, and any practical implications of the findings. Additionally, consider discussing the clinical or psychological significance of the observed effects on eating disorder scores and body comparison scores.

```
[ ]:
```