stat_2

February 18, 2023

[ ]:
```
# import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
```

[ ]:
```
# create the dataframe
df = pd.read_csv('cars_clean.csv')
```

Descriptive Statistics

[ ]:
```
# size of dataframe
df.shape
```

[ ]:
```
# basic statistics
df.describe()
```

[ ]:
```
# count of categorical variables
num_drive_wheels = df['drive-wheels'].value_counts()
print (num_drive_wheels)
```

[ ]:
```
# what are the body styles
df['body-style'].unique()
```

[ ]:
```
# grouping results
grp = df[['drive-wheels', 'body-style', 'price']]
grp_mean = grp.groupby (['drive-wheels', 'body-style'], as_index = False).mean()
grp_mean
```

[ ]:
```
# create a pivot table
grp_pivot = grp_mean.pivot(index = 'drive-wheels', columns = 'body-style')
grp_pivot
```

[ ]:
```
# create a heatmap
plt.pcolor (grp_pivot, cmap = 'RdBu')
plt.colorbar()
plt.show()
```

Pearson Correlation

[ ]:
```
# find correlation between horsepower and price
p_corr, p_val = stats.pearsonr (df['horsepower'], df['price'])
print ('p_corr = ', p_corr, ' ', 'p_value = ', p_val)
```

Compute the Pearson correlation coefficient and the P-value of the following variables against price - wheel-base, length, width, curb-weight, engine-size, bore, city-mpg, highway-mpg

ANOVA (Analysis of Variance)

[ ]:
```
# check the relationship of 'drive-wheels' with 'price'
grp_dw = grp[['drive-wheels', 'price']].groupby(['drive-wheels'])
```

[ ]:
```
# display head
grp_dw.head()
```

[ ]:
```
# get just '4wd' prices
grp_dw.get_group('4wd')['price']
```

[ ]:
```
# use function f_oneway() to run ANOVA
f_val, p_val = stats.f_oneway(grp_dw.get_group('fwd')['price'], grp_dw.
  ↪get_group('rwd')['price'], grp_dw.get_group('4wd')['price'])
print ('f_value = ', f_val, ' ', 'p_value = ', p_val)
```

Examine separately the relationship between: fwd and rwd, rwd and 4wd, 4wd and fwd.