CORRESPONDENCE ANALYSIS

Nupoor Karnik

TO FIND ASSOCIATION BETWEEN CAR MAKE AND HORSEPOWER

CONTINGENCY TABLE

To find the best associated car make and level of horsepower here is the contingency table

Contingency Table				
	High	Low	Medium	Sum
Audi	4	3	12	19
BMW	6	7	7	20
Chevrolet	2	18	7	27
Ford	1	11	11	23
Honda	0	9	6	15
Mercedes	11	5	10	26
Subaru	0	7	4	11
Sum	24	60	57	141

EXPECTED VALUES TABLE

The first thing to calculate is the **expected values** for each cell as if there were no relationships among the two variables.

Chi-Square Statistic Expected Values			
	High	Low	Medium
Audi	3.2340	8.0851	7.6809
BMW	3.4043	8.5106	8.0851
Chevrolet	4.5957	11.4894	10.9149
Ford	3.9149	9.7872	9.2979
Honda	2.5532	6.3830	6.0638
Mercedes	4.4255	11.0638	10.5106
Subaru	1.8723	4.6809	4.4468

OBSERVED MINUS EXPECTED FREQUENCIES

Next, we calculate the difference of the **observed frequencies and the expected frequencies** from the previous tables.

Observed Minus Expected Values			
	High	Low	Medium
Audi	0.76596	-5.08511	4.31915
BMW	2.59574	-1.51064	-1.08511
Chevrolet	-2.59574	6.51064	-3.91489
Ford	-2.91489	1.21277	1.70213
Honda	-2.55319	2.61702	-0.06383
Mercedes	6.57447	-6.06383	-0.51064
Subaru	-1.87234	2.31915	-0.44681

Contributions to the Total Chi-Square Statistic				
	High	Low	Medium	Sum
Audi	0.1814	3.1983	2.4288	5.8084
BMW	1.9793	0.2681	0.1456	2.3930
Chevrolet	1.4661	3.6894	1.4042	6.5596
Ford	2.1703	0.1503	0.3116	2.6322
Honda	2.5532	1.0730	0.0007	3.6268
Mercedes	9.7669	3.3234	0.0248	13.1151
Subaru	1.8723	1.1490	0.0449	3.0663
Sum	19.9895	12.8515	4.3606	37.2016

CONTRIBUTIONS TO CHI SQUARE

Highest chi square value belongs to Mercedes and High horsepower column i.e. 9.766

Lowest chi square belongs to Honda and Medium horsepower column i.e. 0.0007

CHI SQUARE TABLE

Brand vs. Horsepower

The CORRESP Procedure

Inertia and Chi-Square Decomposition

Singular Value	Principal Inertia	Chi-Square	Percent	Cumulative Percent
0.46520	0.21641	30.5145	82.02	82.02
0.21778	0.04743	6.6871	17.98	100.00
	0.26384	37.2016	100.00	

Pr > Chisq = .0002 DF = 12

Summary Statistics for the Row Points			
	Quality	Mass	Inertia
Audi	1.0000	0.1348	0.1561
BMW	1.0000	0.1418	0.0643
Chevrolet	1.0000	0.1915	0.1763
Ford	1.0000	0.1631	0.0708
Honda	1.0000	0.1064	0.0975
Mercedes	1.0000	0.1844	0.3525
Subaru	1.0000	0.0780	0.0824

ROW- MASS INERTIA QUALITY

In the **row** points, notice that the largest **mass** (highest frequency data) is contributed by **Chevrolet** row and the smallest by **Subaru** row.

The largest contribution to the chi square according to the **inertia** metric is **Mercedes** while the lowest in contribution to chi-square total is the **BMW** row.

For **quality** it can be observed that they are all quite close. Representation of data is best for all. Enough explanatory power.

COLUMN- MASS INERTIA QUALITY

Summary Statistics for the Column Points			
	Quality	Mass	Inertia
High	1.0000	0.1702	0.5373
Low	1.0000	0.4255	0.3455
Medium	1.0000	0.4043	0.1172

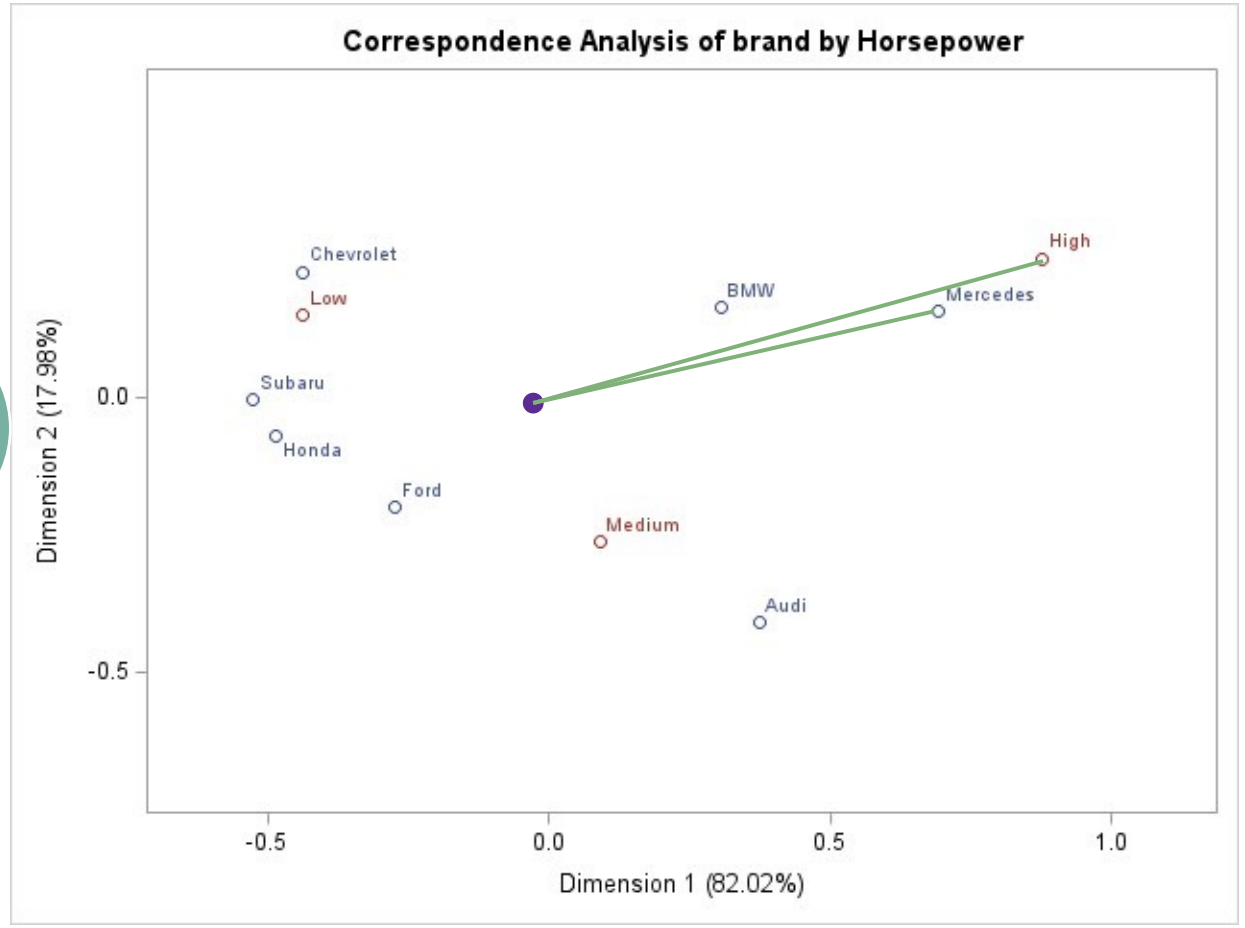
In the column contributions, notice that the largest **mass** (highest frequency data) is contributed by **Low horsepower** and the smallest by **High horsepower**.

The largest contribution to the chi square according to the **inertia** metric is **High horsepower** while the lowest in contribution to chi-square total is the **Medium horsepower**.

For **quality** they are all quite close. Representation of data is best for all.

CORRESPONDENCE MAP

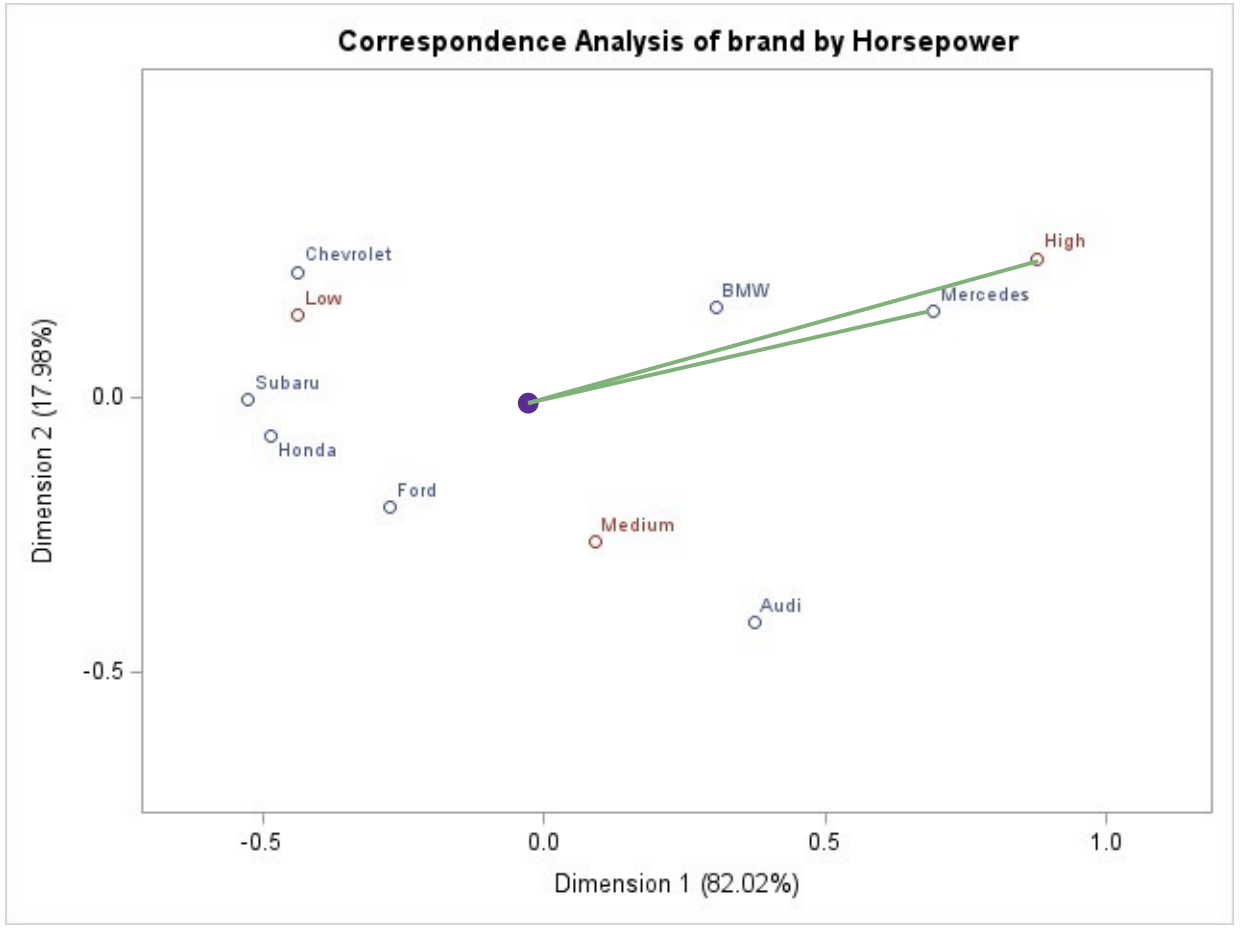
Mercedes + High = Strong Association



CORRESPONDENCE MAP

Dimension 1 — Explains 82.02% of variance

Dimension 2 — Explains 17.98% of variance



Contributions to the Total Chi-Square Statistic				
	High	Low	Medium	Sum
Audi	0.1814	3.1983	2.4288	5.8084
BMW	1.9793	0.2681	0.1456	2.3930
Chevrolet	1.4661	3.6894	1.4042	6.5596
Ford	2.1703	0.1503	0.3116	2.6322
Honda	2.5532	1.0730	0.0007	3.6268
Mercedes	9.7669	3.3234	0.0248	13.1151
Subaru	1.8723	1.1490	0.0449	3.0663
Sum	19.9895	12.8515	4.3606	37.2016

Degrees of Freedom	Area to the Right of Critical Value						
	0.995	0.99	0.975	0.95	0.90	0.10	0.05
1	—	—	0.001	0.004	0.016	2.706	3.841
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026

SIGNIFICANT CHI SQUARE

- Total Chi Sq = 37.20

DF (7-1)*(3-1) = 12

- Chi sq table critical value for DF 12(0.05) = 21.026

Null hypothesis is rejected. There is a relationship between the variable 'Make' and 'Horsepower'.

FREQUENCY TABLE

Positive

- Highest standard residuals(>2) are for Mercedes and high hp
- Audi and Medium hp
- Chevrolet and low hp
- Major source of total chi sq significance results

Negative

- Highest standard residuals for Audi and Low hp
- Mercedes and Low hp
- Major source of total chi sq significance results

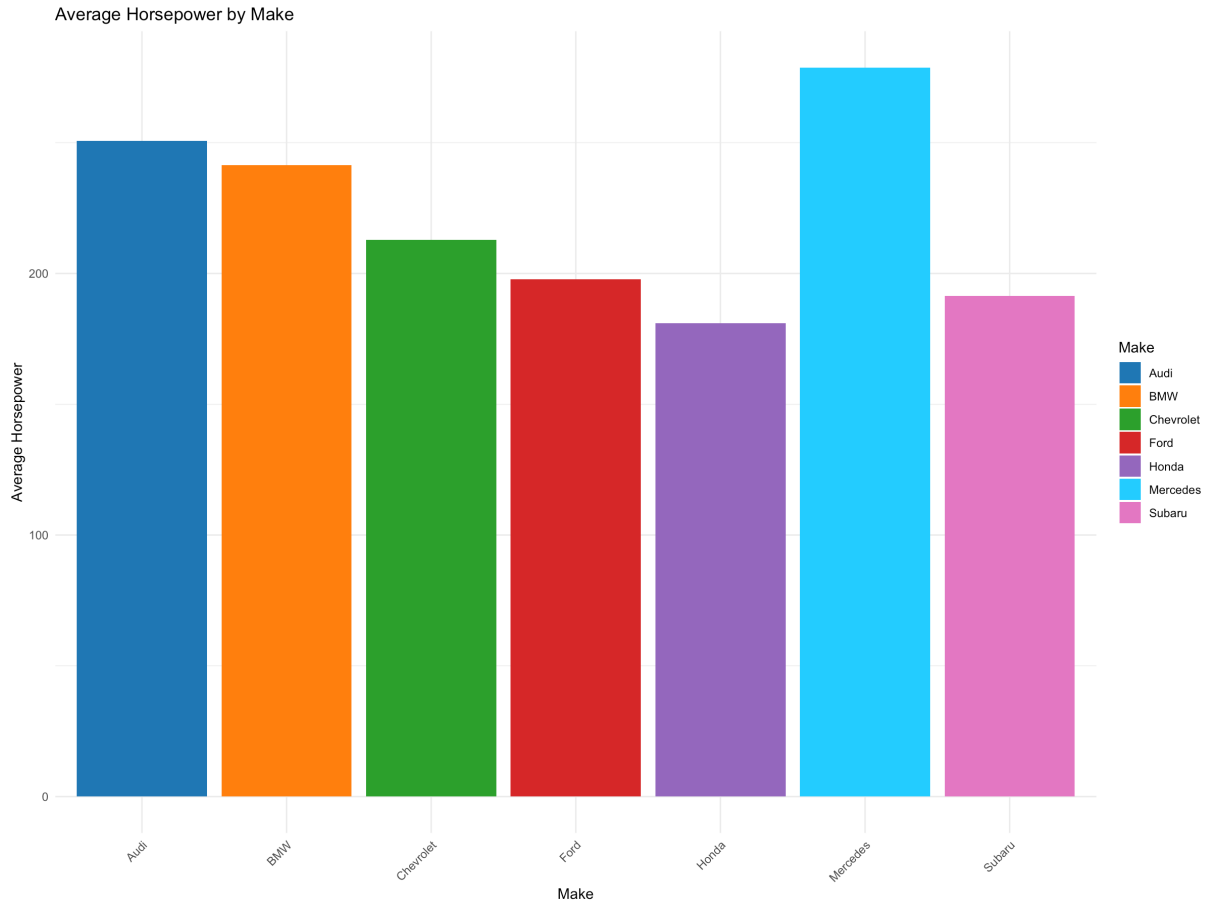
Brand vs. Horsepower

The FREQ Procedure

Table of brand by Horsepower

brand	Horsepower	Frequency	Std Residual	Percent	Row Percent	Column Percent
Audi	Low	3	-2.5366	2.13	15.79	5.00
	Medium	12	2.1707	8.51	63.16	21.05
	High	4	0.5027	2.84	21.05	16.67
	Total	19		13.48	100.00	
BMW	Low	7	-0.7375	4.96	35.00	11.67
	Medium	7	-0.5337	4.96	35.00	12.28
	High	6	1.6672	4.26	30.00	25.00
	Total	20		14.18	100.00	
Chevrolet	Low	18	2.8184	12.77	66.67	30.00
	Medium	7	-1.7074	4.96	25.93	12.28
	High	2	-1.4783	1.42	7.41	8.33
	Total	27		19.15	100.00	
Ford	Low	11	0.5591	7.80	47.83	18.33
	Medium	11	0.7906	7.80	47.83	19.30
	High	1	-1.7679	0.71	4.35	4.17
	Total	23		16.31	100.00	
Honda	Low	9	1.4457	6.38	60.00	15.00
	Medium	6	-0.0355	4.26	40.00	10.53
	High	0	-1.8556	0.00	0.00	0.00
	Total	15		10.64	100.00	
Mercedes	Low	5	-2.6633	3.55	19.23	8.33
	Medium	10	-0.2260	7.09	38.46	17.54
	High	11	3.7989	7.80	42.31	45.83
	Total	26		18.44	100.00	
Subaru	Low	7	1.4729	4.96	63.64	11.67
	Medium	4	-0.2859	2.84	36.36	7.02
	High	0	-1.5644	0.00	0.00	0.00
	Total	11		7.80	100.00	
Total	Low	60		42.55		100.00
	Medium	57		40.43		100.00
	High	24		17.02		100.00
	Total	141		100.00		

BAR GRAPH FOR BRAND VS HORSEPOWER



THANK YOU!