EXERCISE 1a

The FREQ Procedure

Frequency Expected

Table of serumyn by outcome			
	outcome		
serumyn	normal	improved	Total
N	20 13.419	32 38.581	52
Y	12 18.581	60 53.419	72
Total	32	92	124

Since the expected values on the diagonal is more than the observed values & the expectedvalues are less than the observed values in off diagonal cells, there is an association between the variables.
EXERCISE 1b

The FREQ Procedure

Statistics for Table of serumyn by outcome

Statistic	DF	Value	Prob
Chi-Square	1	7.4908	0.0062
Likelihood Ratio Chi-Square	1	7.4399	0.0064
Continuity Adj. Chi-Square	1	6.3957	0.0114
Mantel-Haenszel Chi-Square	1	7.4304	0.0064
Phi Coefficient		0.2458	
Contingency Coefficient		0.2387	
Cramer's V		0.2458	

The p value for the chi-sq test < 0.05 so we reject the null hypothesis. Therefore, there is significant association between both the variables. The p - value for MH test is also < 0.05 so we again reject the null hypothesis. Therefore, there is a linear trend between the variables.
EXERCISE 1c

The FREQ Procedure

Statistics for Table of serumyn by outcome

Column 2 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.6154	0.0675	0.4832	0.7476	0.4702	0.7470
Row 2	0.8333	0.0439	0.7473	0.9194	0.7270	0.9108
Total	0.7419	0.0393	0.6649	0.8190	0.6557	0.8163
Difference	-0.2179	0.0805	-0.3757	-0.0602		
Difference is (Row 1 - Row 2)						

Since the difference (Row 1 – Row 2) is negative and its 95% CI does not include 0, serum users (Row 2 = Y) have a significantly higher probability of improved skin.
EXERCISE 2a

The FREQ Procedure

Frequency Expected Percent

Table of originname by mpg_cat				
	mpg_cat			
originname	inefficient	average	efficient	Total
US	147 100 37.50	52 60 13.27	46 85 11.73	245 62.50
Japan	5 32.245 1.28	17 19.347 4.34	57 27.408 14.54	79 20.15
Europe	8 27.755 2.04	27 16.653 6.89	33 23.592 8.42	68 17.35
Total	160 40.82	96 24.49	136 34.69	392 100.00

Statistics for Table of originname by mpg_cat

Statistic	DF	Value	Prob
Chi-Square	4	120.5468	<.0001
Likelihood Ratio Chi-Square	4	130.6731	<.0001
Mantel-Haenszel Chi-Square	1	75.5390	<.0001
Phi Coefficient		0.5545	
Contingency Coefficient		0.4850	
Cramer's V		0.3921	

Sample Size = 392

Origin and MPG category are strongly associated (Pearson X^2=120.55, $p < 0.0001$; Likelihood-ratio X^2 $p < 0.0001$).
EXERCISE 2b

The FREQ Procedure

Frequency Expected

Table of originname by mpg_cat			
	mpg_cat		
originname	efficient	inefficient	Total
Japan	57 54.175	5 7.8252	62
Europe	33 35.825	8 5.1748	41
Total	90	13	103

Statistics for Table of originname by mpg_cat

Statistic	DF	Value	Prob
Chi-Square	1	2.9327	0.0868
Likelihood Ratio Chi-Square	1	2.8650	0.0905
Continuity Adj. Chi-Square	1	1.9865	0.1587
Mantel-Haenszel Chi-Square	1	2.9042	0.0883
Phi Coefficient		0.1687	
Contingency Coefficient		0.1664	
Cramer's V		0.1687	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	57
Left-sided Pr <= F	0.9772
Right-sided Pr >= F	0.0806
Table Probability (P)	0.0578
Two-sided Pr <= P	0.1286

Sample Size = 103

There is a strong association in the 3*3 table. When we restrict to Japan vs Europe and drop US, the relationship weakens and is not significant.
EXERCISE 2c

The FREQ Procedure

Frequency Percent Row Pct Col Pct

Table of originname by mpg_cat			
	mpg_cat		
originname	efficient	inefficient	Total
Europe	33 32.04 80.49 36.67	8 7.77 19.51 61.54	41 39.81
Japan	57 55.34 91.94 63.33	5 4.85 8.06 38.46	62 60.19
Total	90 87.38	13 12.62	103 100.00

Statistics for Table of originname by mpg_cat

Column 1 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.8049	0.0619	0.6836	0.9262	0.6513	0.9118
Row 2	0.9194	0.0346	0.8516	0.9871	0.8217	0.9733
Total	0.8738	0.0327	0.8097	0.9379	0.7938	0.9311
Difference	-0.1145	0.0709	-0.2534	0.0245		
Difference is (Row 1 - Row 2)						

Column 2 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.1951	0.0619	0.0738	0.3164	0.0882	0.3487
Row 2	0.0806	0.0346	0.0129	0.1484	0.0267	0.1783
Total	0.1262	0.0327	0.0621	0.1903	0.0689	0.2062
Difference	0.1145	0.0709	-0.0245	0.2534		
Difference is (Row 1 - Row 2)						

Sample Size = 103

Because the CI includes 0, the difference is not statistically significant. So we cannot conclude that European cars have a higher probability of high fuel efficiency.
The point estimate suggests Japan is higher.
EXERCISE 3a

The ANOVA Procedure

Class Level Information		
Class	**Levels**	**Values**
originname	3	Europe Japan US

Number of Observations Read	392
Number of Observations Used	392

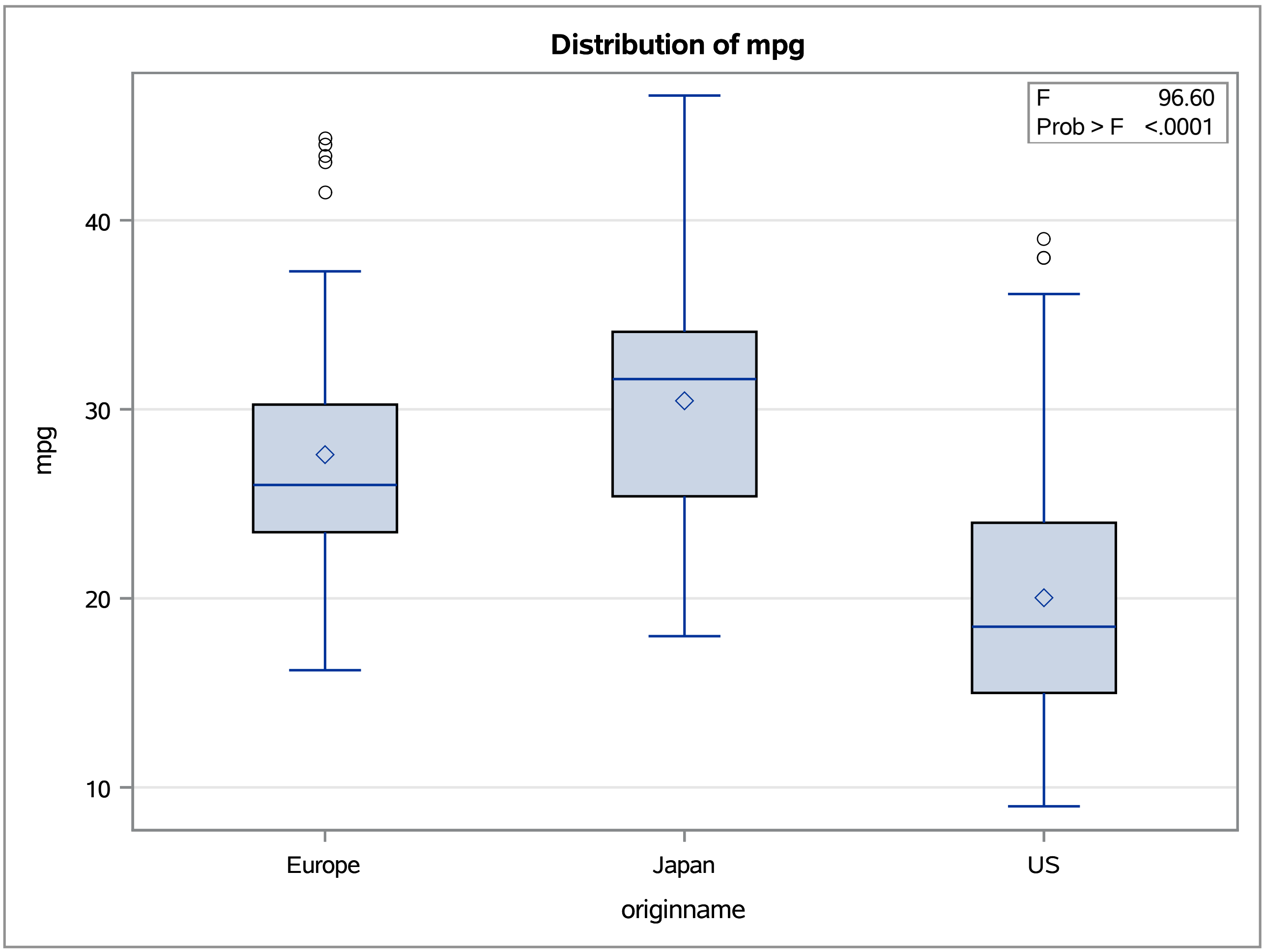
The ANOVA Procedure

Dependent Variable: mpg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	7904.29104	3952.14552	96.60	<.0001
Error	389	15914.70243	40.91183		
Corrected Total	391	23818.99347			

R-Square	Coeff Var	Root MSE	mpg Mean
0.331848	27.28081	6.396236	23.44592

Source	DF	Anova SS	Mean Square	F Value	Pr > F
originname	2	7904.291038	3952.145519	96.60	<.0001

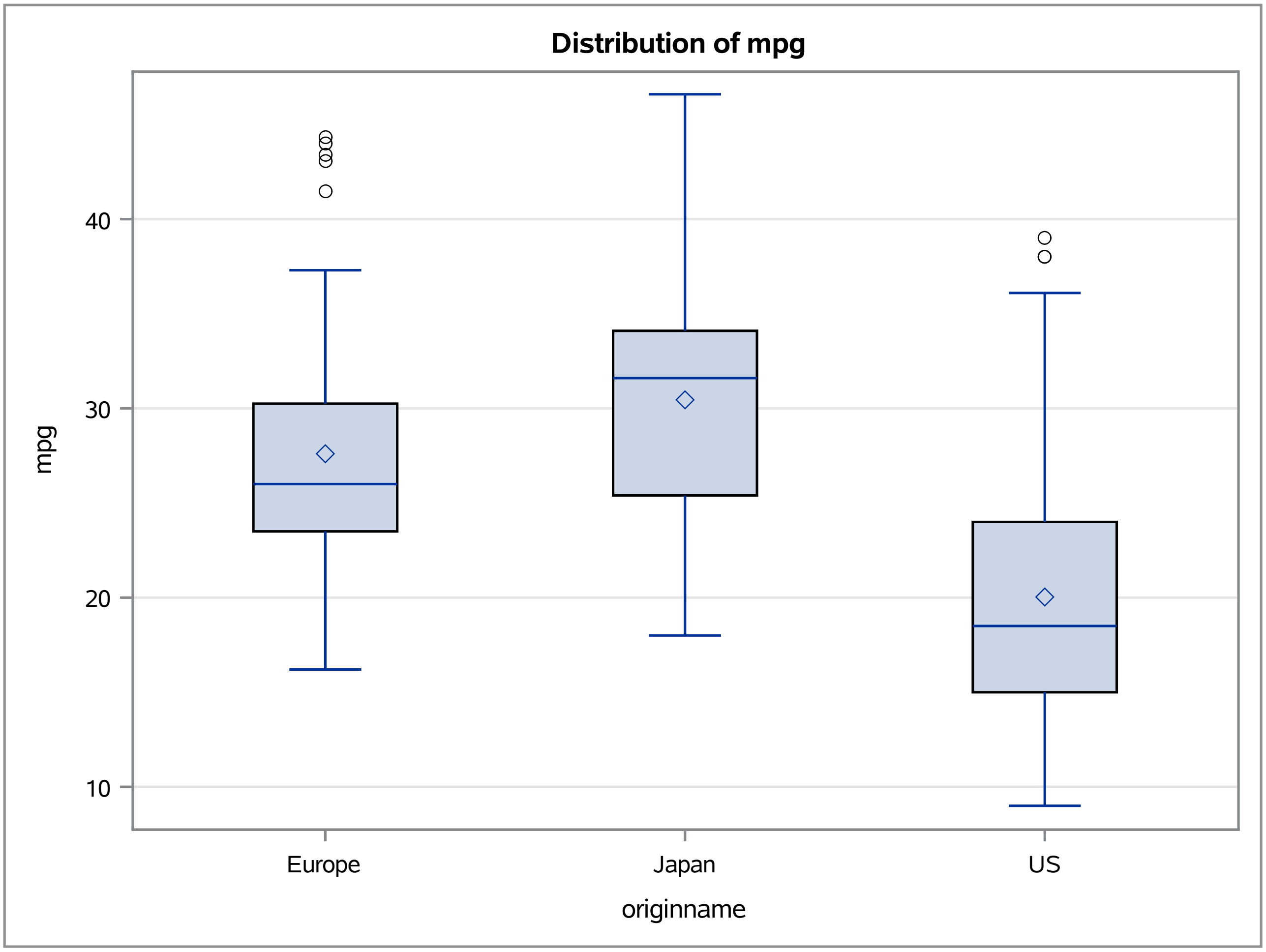


The ANOVA Procedure

Levene's Test for Homogeneity of mpg Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
originname	2	1664.3	832.1	0.24	0.7846
Error	389	1333334	3427.6		

The ANOVA Procedure



Level of originname	N	mpg	
		Mean	Std Dev
Europe	68	27.6029412	6.58018215
Japan	79	30.4506329	6.09004807
US	245	20.0334694	6.44038410

One-way ANOVA of mpg by originname is highly significant ($F = 96.60$, $p < 0.0001$). So mean mpg differs by origin.Levene's test $p = 0.7846$. Group variances look equal, so the standard ANOVA is appropriate.$R^2 = 0.332$ so about 33% of the variation in mpg is explained by origin.All pairs differ, Japan has the highest mpg, US the lowest, Europe in between.
EXERCISE 3b

The ANOVA Procedure

Class Level Information		
Class	**Levels**	**Values**
originname	3	Europe Japan US

Number of Observations Read	392
Number of Observations Used	392

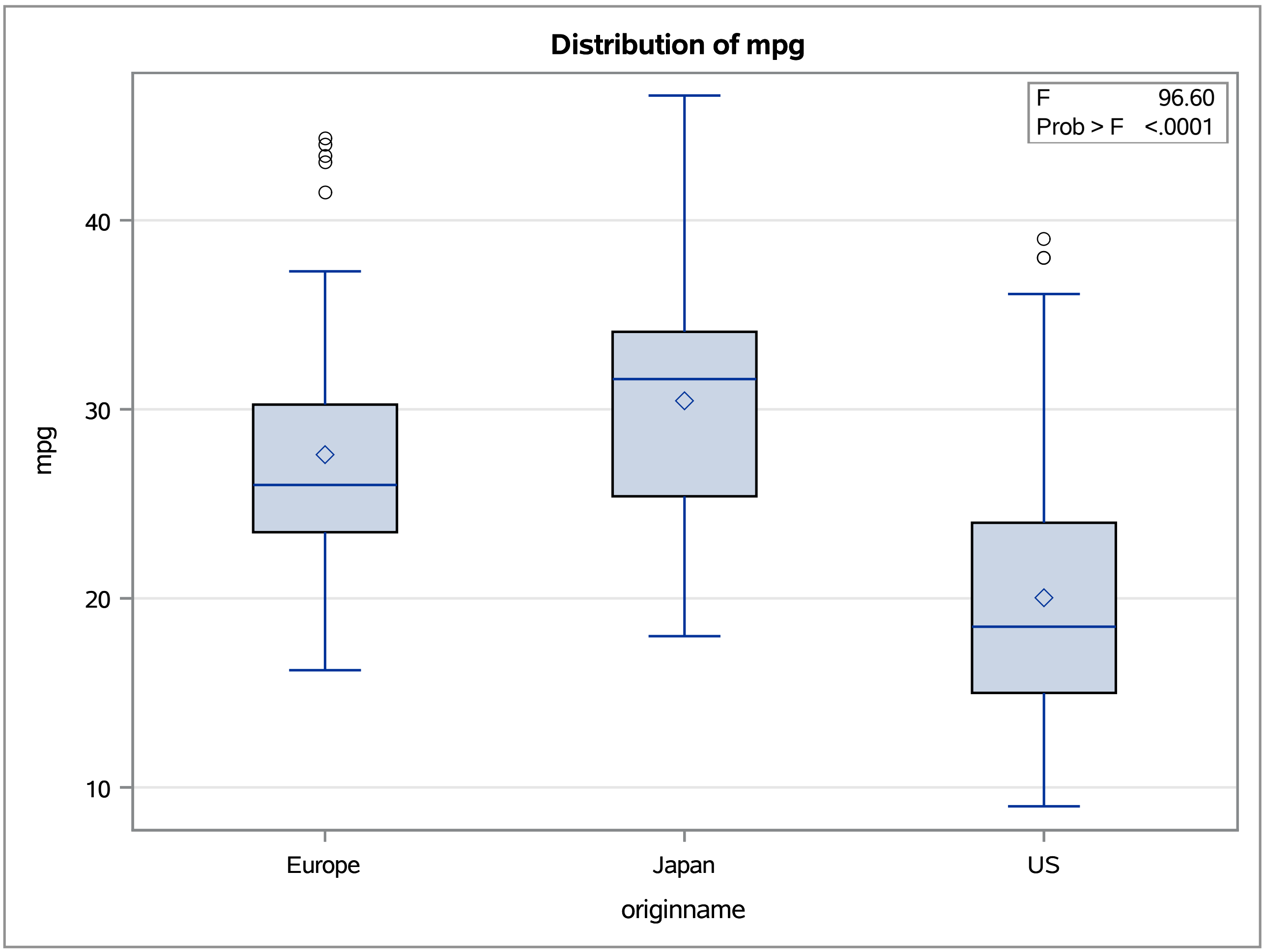
The ANOVA Procedure

Dependent Variable: mpg

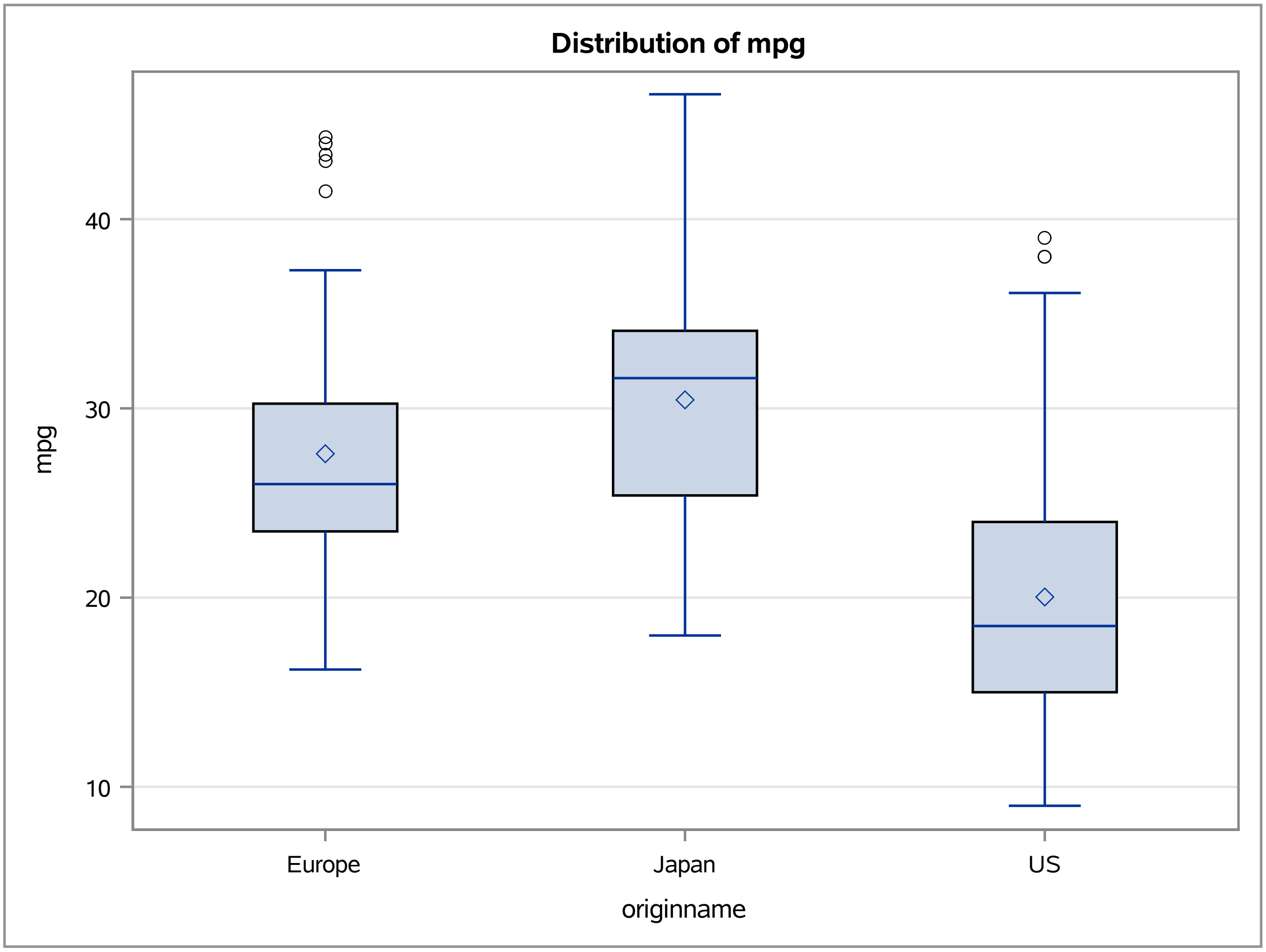
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	7904.29104	3952.14552	96.60	<.0001
Error	389	15914.70243	40.91183		
Corrected Total	391	23818.99347			

R-Square	Coeff Var	Root MSE	mpg Mean
0.331848	27.28081	6.396236	23.44592

Source	DF	Anova SS	Mean Square	F Value	Pr > F
originname	2	7904.291038	3952.145519	96.60	<.0001



The ANOVA Procedure



The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for mpg

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	389
Error Mean Square	40.91183
Critical Value of Studentized Range	3.32725

Comparisons significant at the 0.05 level are indicated by ***.				
originname Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
Japan - Europe	2.8477	0.3583	5.3370	***
Japan - US	10.4172	8.4701	12.3642	***
Europe - Japan	-2.8477	-5.3370	-0.3583	***
Europe - US	7.5695	5.5068	9.6321	***
US - Japan	-10.4172	-12.3642	-8.4701	***
US - Europe	-7.5695	-9.6321	-5.5068	***

All CIs exclude 0, so all three origins differ significantly. Cars from Japan have thehighest fuel efficiency, Europe is intermediate, and the US has the lowest mpg. The gaps are largest between Japan and the US (~10 mpg) and moderate between Europe and Japan (~3 mpg).
EXERCISE 4a

The ANOVA Procedure

Class Level Information		
Class	**Levels**	**Values**
mpg_cat	3	average efficient inefficient

Number of Observations Read	392
Number of Observations Used	392

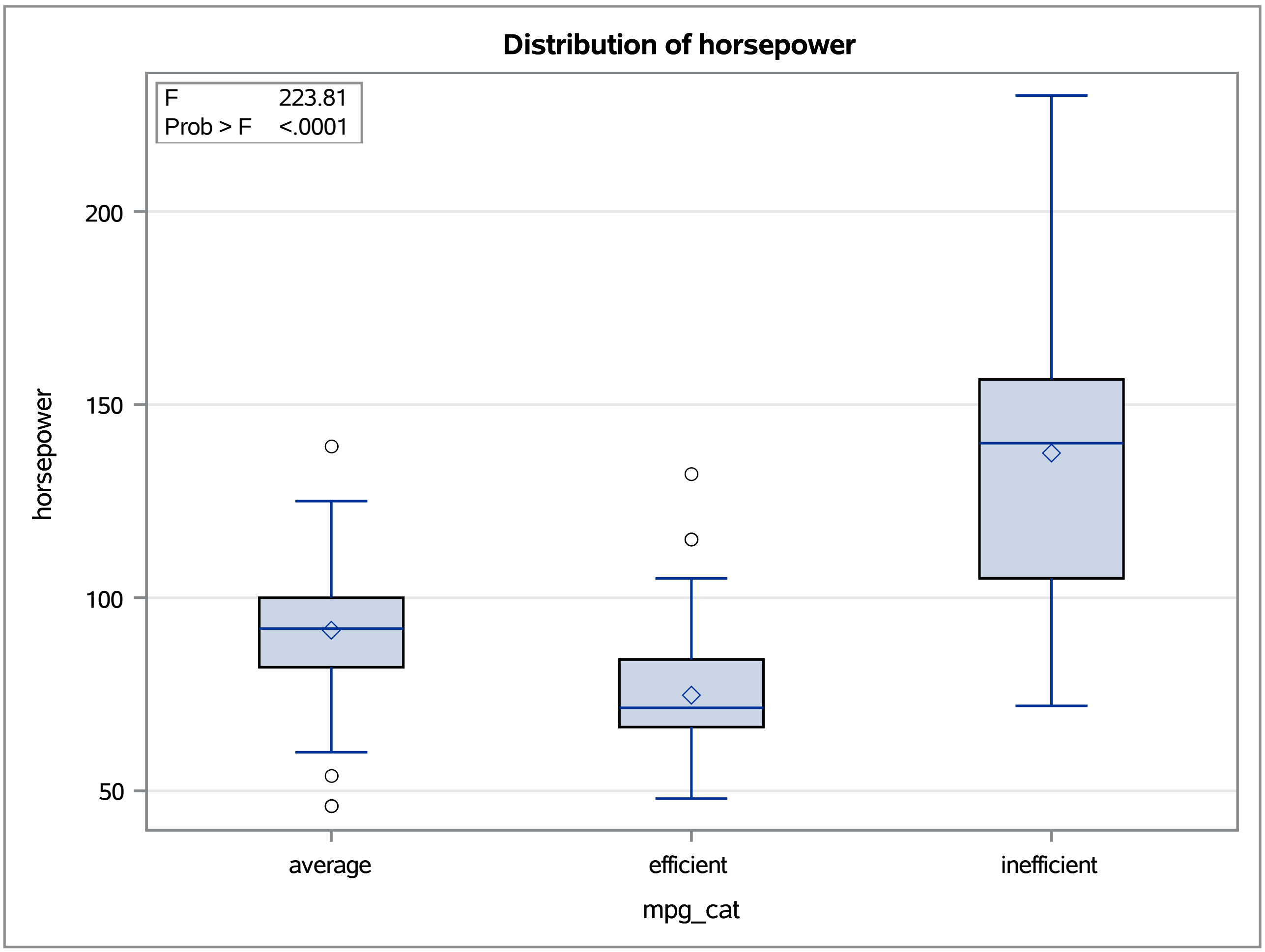
The ANOVA Procedure

Dependent Variable: horsepower

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	309944.6092	154972.3046	223.81	<.0001
Error	389	269349.0234	692.4139		
Corrected Total	391	579293.6327			

R-Square	Coeff Var	Root MSE	horsepower Mean
0.535039	25.18801	26.31376	104.4694

Source	DF	Anova SS	Mean Square	F Value	Pr > F
mpg_cat	2	309944.6092	154972.3046	223.81	<.0001



The ANOVA Procedure

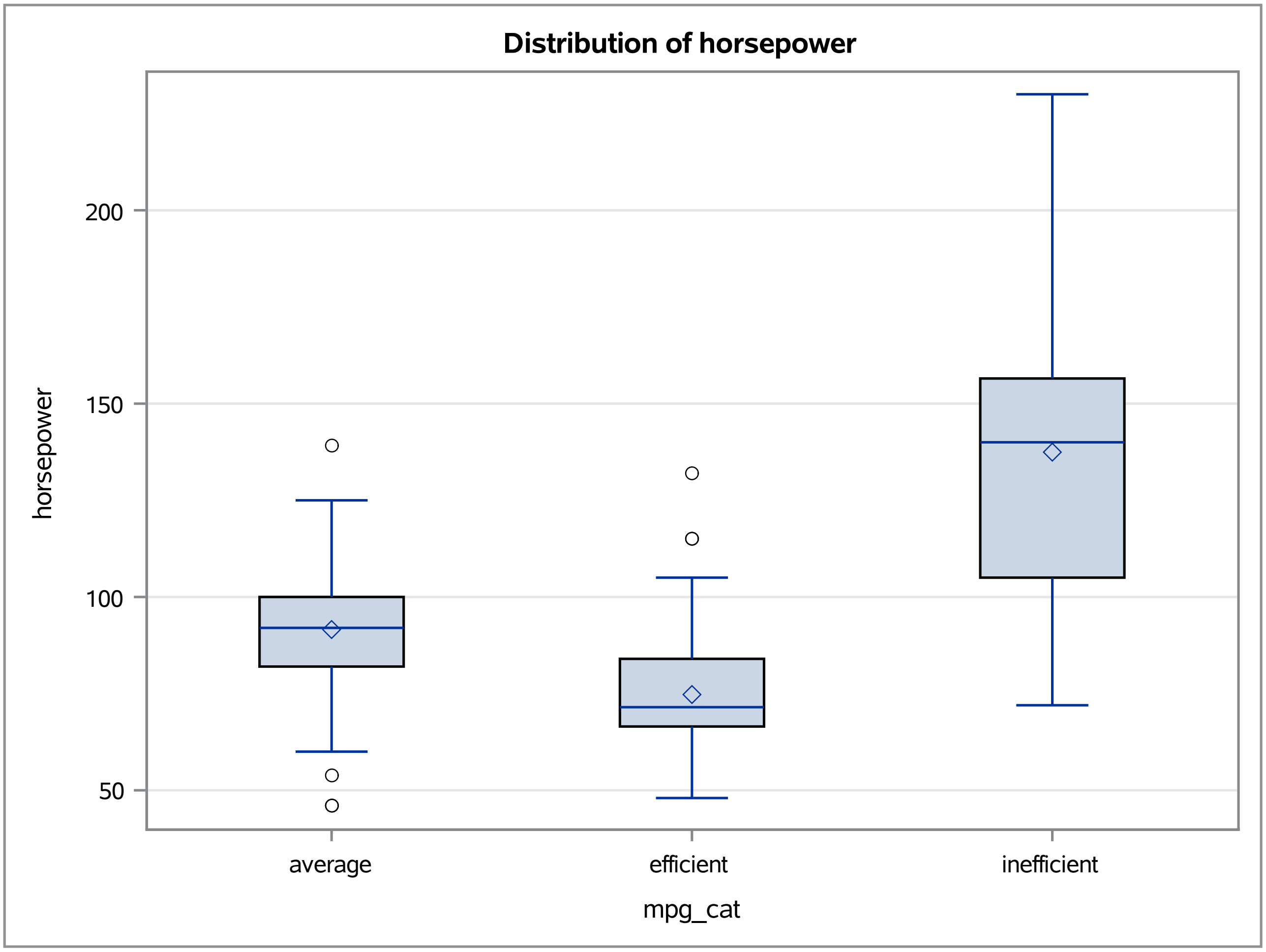
Levene's Test for Homogeneity of horsepower Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
mpg_cat	2	1.2487E8	62434202	48.36	<.0001
Error	389	5.0223E8	1291082		

Welch's ANOVA for horsepower

Source	DF	F Value	Pr > F
mpg_cat	2.0000	203.00	<.0001
Error	235.9		

The ANOVA Procedure



The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for horsepower

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	389
Error Mean Square	692.4139
Critical Value of Studentized Range	3.32725

Comparisons significant at the 0.05 level are indicated by ***.				
mpg_cat Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
inefficient - average	45.877	37.885	53.869	***
inefficient - efficient	62.678	55.457	69.898	***
average - inefficient	-45.877	-53.869	-37.885	***
average - efficient	16.801	8.548	25.053	***
efficient - inefficient	-62.678	-69.898	-55.457	***
efficient - average	-16.801	-25.053	-8.548	***

One-way ANOVA of horsepower by mpg_cat is highly significant ($F = 223.81$, $p < 0.0001$). So mean horsepower differs by mpg_cat.Levene’s test p = <.0001. Group variances look unequal, so the standard ANOVA is inappropriate.Welch test p = <.0001.$R^2 = 0.535039$ so about 53% of the variation in horsepower is explained by mpg_cat.All pairs differ, Japan has the highest horsepower, US the lowest, Europe in between.Inefficient cars have the highest horsepower, efficient the lowest, with average in between.
EXERCISE 4b

The ANOVA Procedure

Class Level Information		
Class	**Levels**	**Values**
mpg_cat	3	average efficient inefficient

Number of Observations Read	392
Number of Observations Used	392

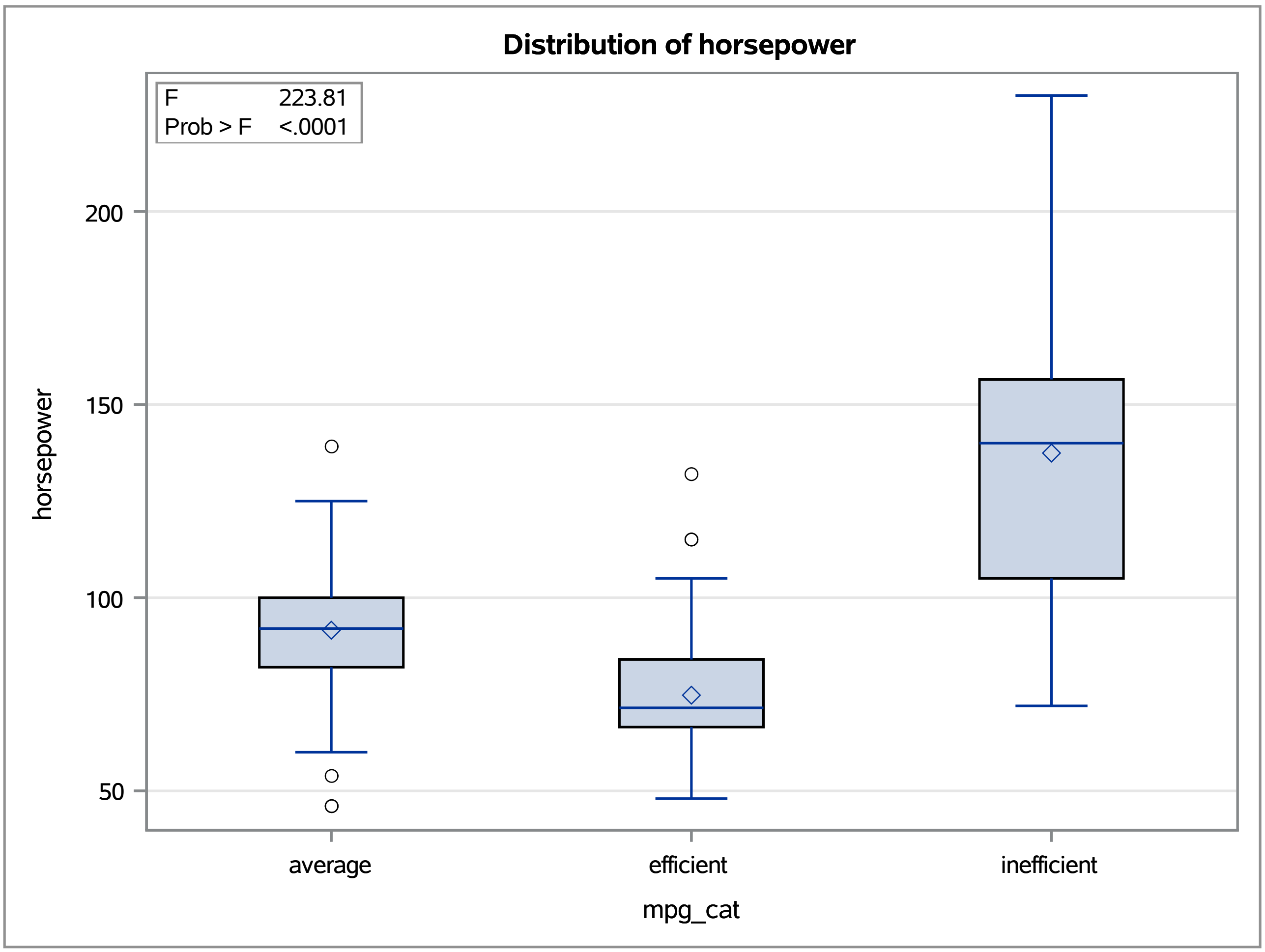
The ANOVA Procedure

Dependent Variable: horsepower

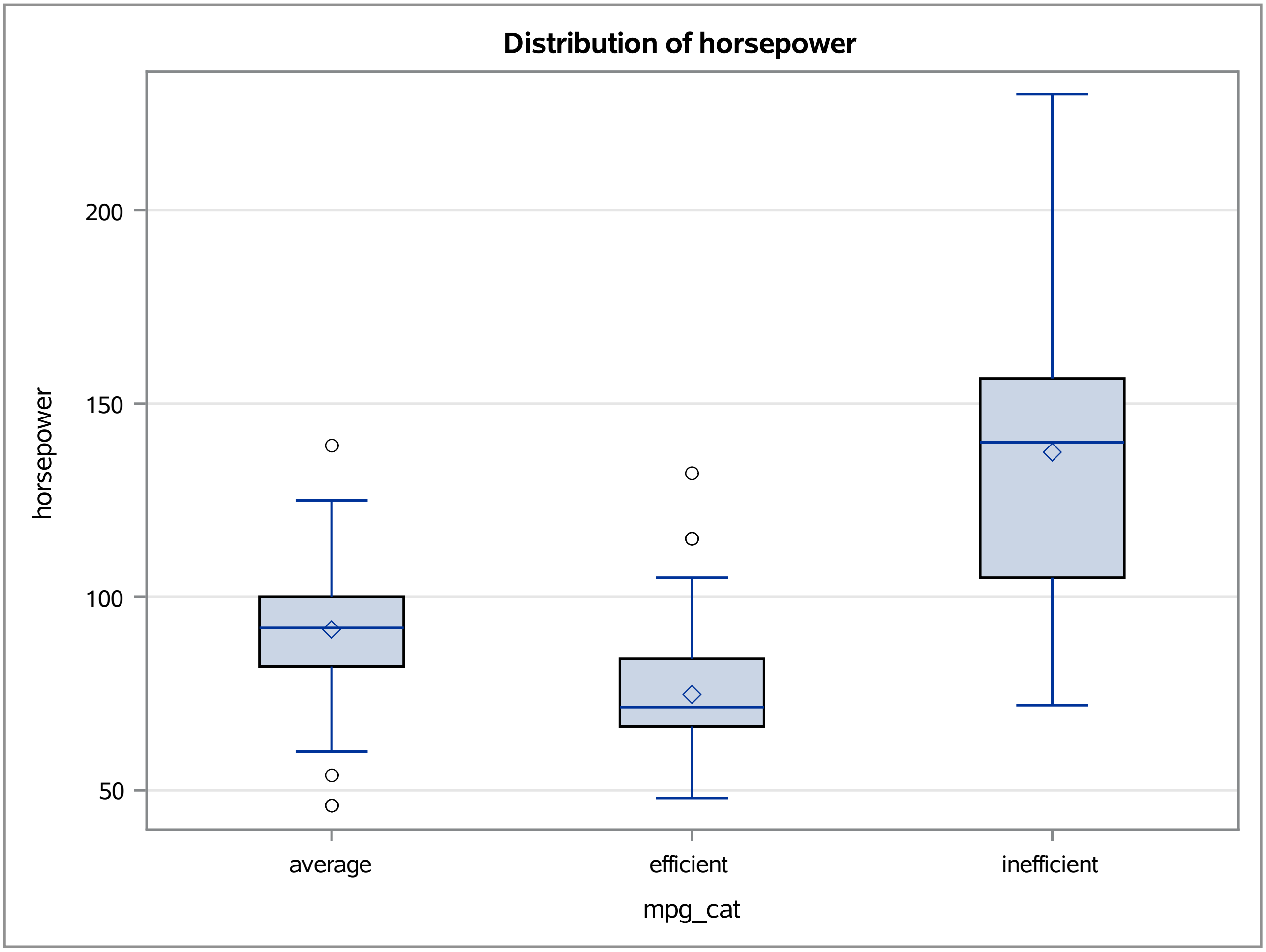
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	309944.6092	154972.3046	223.81	<.0001
Error	389	269349.0234	692.4139		
Corrected Total	391	579293.6327			

R-Square	Coeff Var	Root MSE	horsepower Mean
0.535039	25.18801	26.31376	104.4694

Source	DF	Anova SS	Mean Square	F Value	Pr > F
mpg_cat	2	309944.6092	154972.3046	223.81	<.0001



The ANOVA Procedure



The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for horsepower

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	389
Error Mean Square	692.4139
Critical Value of Studentized Range	3.32725

Comparisons significant at the 0.05 level are indicated by ***.				
mpg_cat Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
inefficient - average	45.877	37.885	53.869	***
inefficient - efficient	62.678	55.457	69.898	***
average - inefficient	-45.877	-53.869	-37.885	***
average - efficient	16.801	8.548	25.053	***
efficient - inefficient	-62.678	-69.898	-55.457	***
efficient - average	-16.801	-25.053	-8.548	***

All CIs exclude 0, so all three origins differ significantly. Horsepower decreases as fuel efficiency improves: Inefficient > Average > Efficient.