UC SAN DIEGO EXTENSION
CSE-41069 INTRODUCTION TO STATISTICS

RESEARCH AND FINDINGS ON THE RELATIONSHIP BETWEEN HORSEPOWER AND ENGINE SIZE

By

GROUP 2
KANG SANG YUN
PARK WOO JIN
RACHEL LEE WEI YEN
TIMOTHY CHAY

Table of Contents

List of Tables ..... **iii**

List of Figures ..... **iii**

List of Appendix ..... **iv**

Introduction **1**

Methodology and Tools **2**

1. Methodology *2*

2. Tools *2*

Key Questions **3**

Statistical Analysis **4**

1. Is there a significant difference in Horsepower between cars with medium and large engines? *4*

Test Hypotheses Regarding the Difference of Two Independent Means 4

Test Requirements 4

Verifying the Data Distribution 4

Calculating the Test Statistic 6

Test Conclusion and Interpretation Statement 9

Constructing a (1- α) • 100% Confidence Interval for the Difference of Two Means 9

Interpretation Statement 9

2. Is there an association between horsepower and engine size? (Use four categories for HP) *11*

Test Requirements 11

Verifying the Expected Frequencies (test requirements) 11

Determining the Expected Proportion 12

Determining the Expected Count (Expected Frequencies) 12

Calculating the Test Statistic 12

Test Conclusion and Interpretation Statement 13

3. Is there a relationship between car horsepower and engine size? *15*

Hypothesis Test Regarding the Slope Coefficient, β_1 15

Test Requirements 15

Verifying the Data Distribution 15

Calculating the Test Statistic 17

Test Conclusion and Interpretation Statement 19

Constructing a (1- α) • 100% Confidence Interval for the Slope of the Regression Line 20
Interpretation Statement 20
Conclusion 21
Lessons Learned 22
Reference 23
Appendix 24
Appendix I: Original cars dataset 24
Appendix II: Critical Values for Student's *t*-Distributions 25
Appendix III: Chi-Square Distribution 26
Appendix IV: Normal probability data for horsepower of cars with medium engines 27
Appendix V: Normal probability data for horsepower of cars with large engines 28
Appendix VI: Summary table of $O_i - E_i$ 29
Appendix VII: Summary table of $(O_i - E_i)^2$ 29
Appendix VIII: Summary table of $(O_i - E_i)^2 / E_i$ 29
Appendix IX: Summary of residual calculation between horsepower and engine size 30
Appendix X: R code snippet for Q1 32
Appendix XI: R code snippet for Q2 32
Appendix XII: R code snippet for Q3 32

List of Tables

Table 1.1: Horsepower of cars with medium and large engines 4
Table 1.2: Data summary for horsepower between cars with large and medium engines 5
Table 1.3: Summary table of data analysis for Q1 7
Table 1.4: Summary of Q1 analysis 9
Table 2.1: Relative marginal frequencies for engine size and horsepower categories 12
Table 2.2: Summary table for expected proportion 12
Table 2.3: Summary table of expected count 12
Table 2.4: Summary of Q2 analysis 14
Table 3.1: Engine size and horsepower data 15
Table 3.2: Summary table of data analysis for Q3 18
Table 3.3: Summary of Q3 analysis 20

List of Figures

Figure 1.1: Normal probability plot for large engine 5
Figure 1.2: Normal probability plot for medium engines 5
Figure 1.3: Boxplots for horsepower 6
Figure 1.4: Test hypotheses for difference between mean of two independent populations 6
Figure 1.5: Illustration on critical region and test statistic 8
Figure 1.6: Illustration of P-value 8
Figure 2.1: Distribution of engine size by horsepower capacity 11
Figure 3.1: Least-squares regression equation on scatter diagram of horsepower and engine size 16
Figure 3.2: Normal probability plot of residuals 17
Figure 3.3: Test hypotheses regarding a slope coefficient 17

List of Appendix

Appendix I: Original cars dataset **24**
Appendix II: Critical Values for Student's *t*-Distributions **25**
Appendix III: Chi-Square Distribution **26**
Appendix IV: Normal probability data for horsepower of cars with medium engines **27**
Appendix V: Normal probability data for horsepower of cars with large engines **28**
Appendix VI: Summary table of $O_i - E_i$ **29**
Appendix VII: Summary table of $(O_i - E_i)^2$ **29**
Appendix VIII: Summary table of $(O_i - E_i)^2 / E_i$ **29**
Appendix IX: Summary of residual calculation between horsepower and engine size **30**
Appendix X: R code snippet for Q1 **32**
Appendix XI: R code snippet for Q2 **32**
Appendix XII: R code snippet for Q3 **32**

Introduction

Up until recently, a car's engine size has always been a strong indicator of a car's performance and power. A car's power is measured by its engine size, or also known as its 'capacity' or 'displacement'. The engine size is expressed in cubic centimeters (cc) and refers to the total volume in which the engine's pistons operate. Typically, the larger this number, the more powerful the car, because of its ability to push more air and fuel through the engine, thus, allowing it to accelerate faster and tow heavier loads.

However, with the rapid advancement in technology in the automotive industry, car manufacturers can now produce cars with more power but smaller engine sizes. This report explores this scenario to understand the effect of technology advancement on performance and the relationship between engine size and horsepower.

Using a set of sample data with 92 cars, containing a variety of categorical and quantitative data, statistical methods, such as **T-statistic hypothesis test**, **Chi-squared test**, and a **Least-Squares regression model,** are employed to investigate any interesting or meaningful relationships for cars in the upcoming model year, particularly between horsepower and engine size. Additionally, this report seeks to find out if the results are statistically significant between different categories of cars. Because we were focused on better defining the relationship between engine size and horsepower, we extracted data from the most relevant variables, including **Engine Size, Engine Size Category**, **Horsepower**, and **Horsepower Category**. The Statistical Analysis section describes in detail the types of tests performed and its interpretation. The section is broken down into three parts (key questions) as follows:

- Is there a significant difference in Horsepower between cars with medium and large engines?
- Is there an association between horsepower and engine size? (Use four categories for HP)
- Is there a relationship between car horsepower and engine size?

A comprehensive coverage of our data transformations will be included in the Appendix section.

Methodology and Tools

1. Methodology

In conducting the research, a variety of statistical methodologies such as T-statistic hypothesis test (for testing the difference in means in horsepower between cars with medium and large engines), Chi-squared test for independence (for determining if there is an association between (categorical) horsepower and engine size), and a Least-Squares regression model (to test whether a linear relationship exists between horsepower and engine size) were applied.

Each member in the group works on all the questions individually, using their own preferred statistical software. The results were then compiled and compared, followed by a group discussion to mutually conclude on the final report result and investigation conclusion, ensuring that each member is able to understand their mistakes and comprehends the final results.

2. Tools

A combination of tools was used to analyze and compare the statistical results to achieve a higher level of accuracy.

Although each member worked individually on the report questions, generally, all data to be analyzed are prepared using Microsoft Excel. The prepared data is then imported in Microsoft Excel or R for further statistical analysis.

Key Questions

Based on initial understanding on the impact of technology on a car's performance and power, the following questions were identified as being most important to address in the analysis:

- Is there a significant difference in Horsepower between cars with medium and large engines?
- Is there an association between horsepower and engine size? (using four categories for horsepower)
- Is there a relationship between car horsepower and engine size?

The three questions were picked for their relevance to the topic of the report, which is to study and relationship between a car's engine size and horsepower.

Statistical Analysis

This section describes the process of analyzing the data with the purpose of answering the key questions defined. The section is broken down into three parts with no order of importance.

1. Is there a significant difference in Horsepower between cars with medium and large engines?

Since horsepower is the amount of power produced by a car's engine, we begin our analysis by checking if there is a significant difference in horsepower between cars with medium and large engines. Specifically, we would like to see if there is any difference in mean values between the samples from the medium sized and large sized engines. In the following discussion, we will refer to each car model as a car, while a medium sized engine is defined as having an engine between 2.0L – 2.99L and a large sized engine as having an engine size of 3.0L and above.

Table 1.1 shows the list of horsepower for cars with medium and large engines respectively based on the cars sample data provided in Appendix I.

Medium Engines	Large Engines
172 140 164 130 150 140 130 134 114 172 128 130 168 160 135 178 138 109 110 110 100 96 155 110 110 110 100 110 93 115 105	200 208 278 185 225 202 160 155 151 141 170 180 200 170 153 147 214 190 210 170 170 170 295 140 160 140 200 160 300 300 160 170 165 142 145 170

Table 1.1: Horsepower of cars with medium and large engines

Test Hypotheses Regarding the Difference of Two Independent Means

To identify if there is a significant difference between the horsepower of cars with medium and large engine, we conduct a hypothesis test to test the difference of the two means.

Test Requirements

- The sample is obtained using random sampling
- Sample has no outliers and the population from which the sample is drawn is normally distributed, or large (more than or equal to 30)
- For each sample, the sample size is no more than 5% of the population size
- The samples are independent

Verifying the Data Distribution

Before we begin our test, we first need to verify that each sample comes from a population that is approximately normal with no outliers by preparing normal probability plots and boxplots.

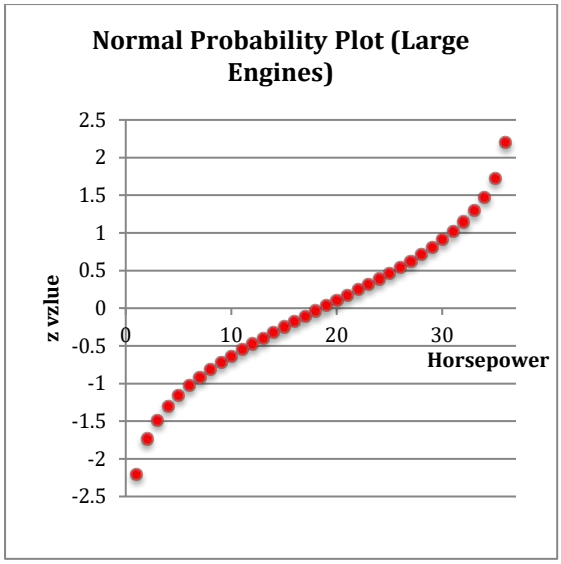


Figure 1.1: Normal probability plot for large engine

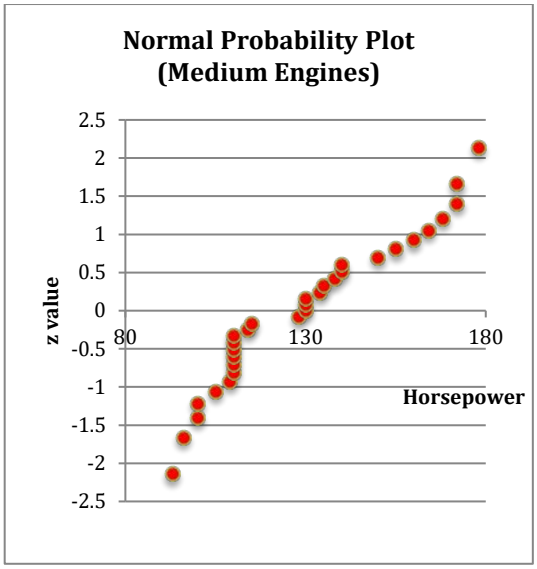


Figure 1.2: Normal probability plot for medium engines

	Large Engines	Medium Engines
Sample Size, *n*	36	31
Mean, $\bar{x}$	186.0	129.5
Standard Deviation, *s*	44.8	25.1
Median	170.0	130.0
First Quartile, Q1	158.8	110.0
Third Quartile, Q3	200.5	145.0
Inter Quartile Range, IQR	41.8	35
Min Value	140.0	93.0
Max Value	300.0	178.0
Lower Fence	96.1	57.5
Upper Fence	263.1	197.5
Outliers	300, 300, 295, 278	-

Table 1.2: Data summary for horsepower between cars with large and medium engines

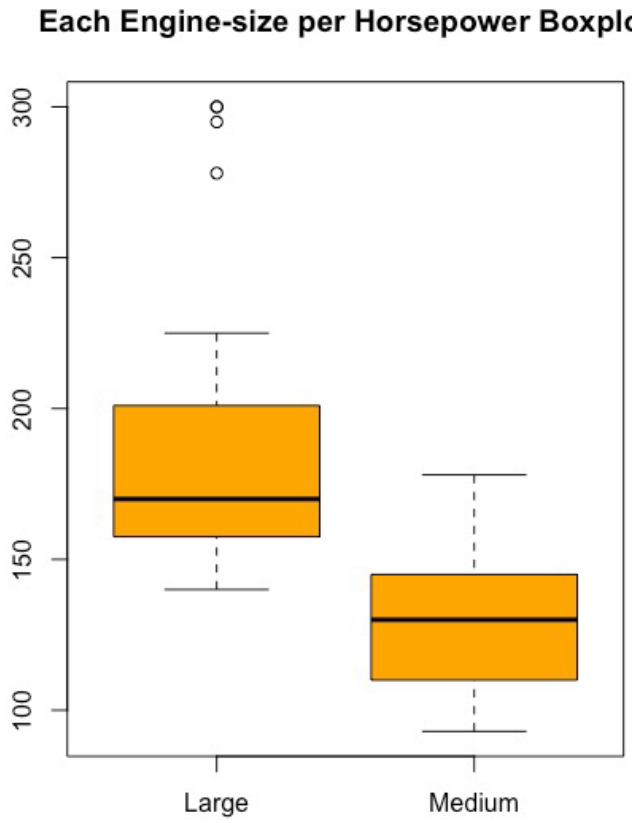


Figure 1.3: Boxplots for horsepower

Based on the normal probability plots in Figure 1.1 and Figure 1.2, as well as the boxplots in Figure 1.3, we are able to determine that the sample data came from populations that are approximately normally distributed. As for the boxplots, it seems there is somewhat of a difference between the horsepower for both large and medium engines.

In addition, the sample data were also obtained through random sampling.

Now that we are able to estimate the sampling distribution of horsepower for cars with large and medium engine size and confirmed that the samples were are randomly selected, we proceed with our hypotheses testing on whether the evidence suggest that the cars of large sized engines and horsepower from the cars with medium engines have a difference that is significant or due to random sampling error.

Calculating the Test Statistic

Having met all the test requirements, we proceed with our test by first determining the null and alternative hypotheses and selecting a level of significance.

Two-Tailed	Left-Tailed	Right-Tailed
$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$	$H_1: \mu_1 < \mu_2$	$H_1: \mu_1 > \mu_2$

Figure 1.4: Test hypotheses for difference between mean of two independent populations

where μ_1 is the population mean for horsepower with large engines
μ_2 is the population mean for horsepower with medium engines

For the purposes of this report, we select a two-tailed hypothesis test and will use a (1- α) % of 95% and a Significance Level, α, of 0.05.

Summary Table	
Null hypothesis	The mean horsepower of cars with medium sized engines is **equal** to the mean horsepower of cars with large engines
Alternative hypothesis	The mean horsepower of cars with medium sized engines is **not equal** to the mean horsepower of cars with large engines
Level of significance, α	0.05
Degree of Freedom	30

Table 1.3: Summary table of data analysis for Q1

Using the **Classical Approach**, we compute the test statistic

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

which approximately follows Student's *t*-distribution.

$$t_0 = \frac{(186 - 129.5) - 0}{\sqrt{\frac{25.1^2}{31} + \frac{44.8^2}{36}}}$$

$$t_0 = 6.472$$

The test statistic is t_0 = 6.472. The point is labeled in Figure 1.5.
This is a two-tailed test with α = 0.05 and 30 degrees of freedom (taken from smaller sample size – 1). Using the table in Appendix II, the critical values with 30 degrees of freedom are $t_{\alpha/2} = t_{0.05/2} = t_{0.025}$ = 2.042 and $-t_{0.025}$ = -2.042. The critical regions are displayed in Figure 1.5.

Because the test statistic t_0 = **6.472** > Critical Value **2.042**, **we reject the hull hypothesis**, H_0.

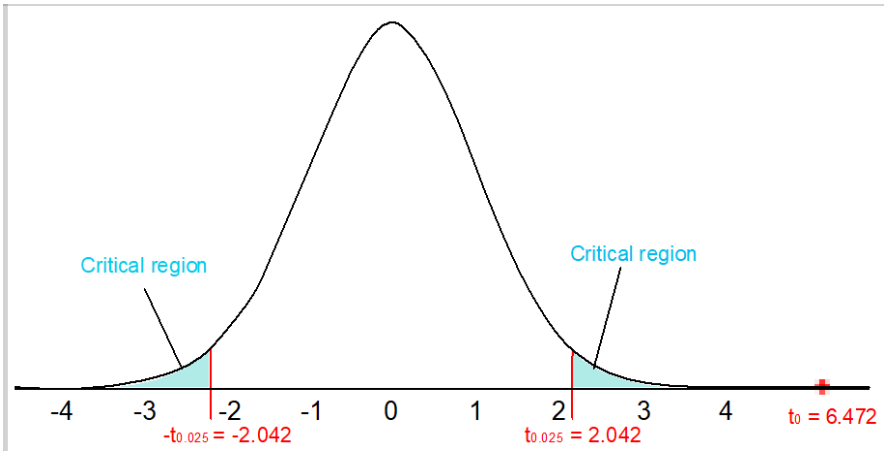


Figure 1.5: Illustration on critical region and test statistic

Using the **P-Value Approach**, we compute the test statistic

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t_0 = \frac{(186 - 129.5) - 0}{\sqrt{\frac{25.1^2}{31} + \frac{44.8^2}{36}}}$$

$$t_0 = 6.472$$

Because this is a two-tailed test, the P-value is the area under the t-distribution to the left of t_0 = -6.472 and to the right of t_0 = 6.472. Refer to Figure 1.6.

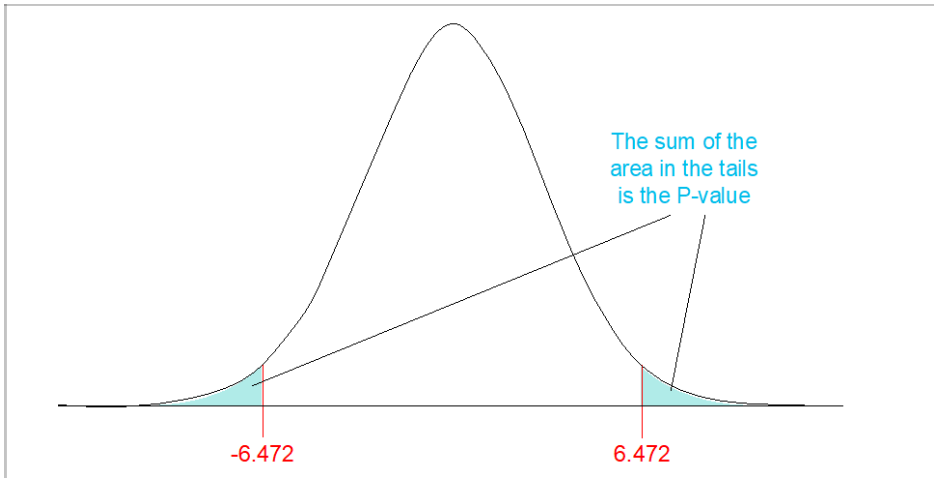


Figure 1.6: Illustration of P-value

Using the table in Appendix II to estimate the area under the *t*-distribution to the right of t_0 = 6.472 and multiply it by two (as graph is symmetrical).

$$\text{P-value} = P(t_0 < -6.472 \; or > 6.472)$$

$$= 2P(t_0 > 6.472)$$

As the value of 6.472 > 3.646, which is the last available value given for 30 degrees of freedom in Appendix II, we assume the P-value to be < 0.0005.

Because the P-value is smaller than the level of significance (0.0005 < 0.05), **we reject the null hypothesis**.

Test Conclusion and Interpretation Statement

The sample data suggest that there is sufficient evidence at the α = 0.05 level of significance to support the statement that there is a significant difference in horsepower between cars with medium and large engines.

Summary Table	
Test Statistic	6.472
Critical Values	2.042, -2.042
Compare critical value with test statistic	6.472 > 2.042
Test conclusion	Reject H0
Interpretation Statement	Since our test-statistic 6.472 > critical value 2.042, there is sufficient evidence at a 95% level of significance to support the statement that there is a significant difference in horsepower between cars with medium and large engines.

Table 1.4: Summary of Q1 analysis

Constructing a (1- α) • 100% Confidence Interval for the Difference of Two Means

Using the data presented in Table 1.4, we construct the confidence level by calculating the Lower and Upper bounds with the α = 0.05 using the formula shown below:

$$Lower\ bound = (\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_1}}$$

$$= (186 - 129.5) - 2.042 \cdot \sqrt{\frac{44.8^2}{36} + \frac{25.1^2}{31}}$$

$$= 38.64$$

$$Upper\ bound = (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_1}}$$

$$= (186 - 129.5) + 2.042 \cdot \sqrt{\frac{44.8^2}{36} + \frac{25.1^2}{31}}$$

$$= 74.26$$

Interpretation Statement

We are 95% confident the mean difference between horsepower of cars with large engines and horsepower of cars with medium engines is between 38.64 and 74.26. The confidence interval does not contain a zero value. Therefore, there is sufficient evidence to conclude there is a difference in the

horsepower of cars with large and horsepower of cars with medium engines. That is, the engine size does seem to affect the car's horsepower significantly.

2. Is there an association between horsepower and engine size? (Use four categories for HP)

To understand if there is an association between horsepower and engine size, we will conduct a **Chi-square Test for Independence**, as this is an inferential method to determine whether there is an association between 2 categorical variables.

Test Requirements

- All expected frequencies are greater than or equal to 1 (all $E_i \geq 1$)
- No more than 20% of the expected frequencies are less than 5

Verifying the Expected Frequencies (test requirements)

Figure 2.1 shows the observed frequencies of engine size by horsepower category as provided in the sample data.

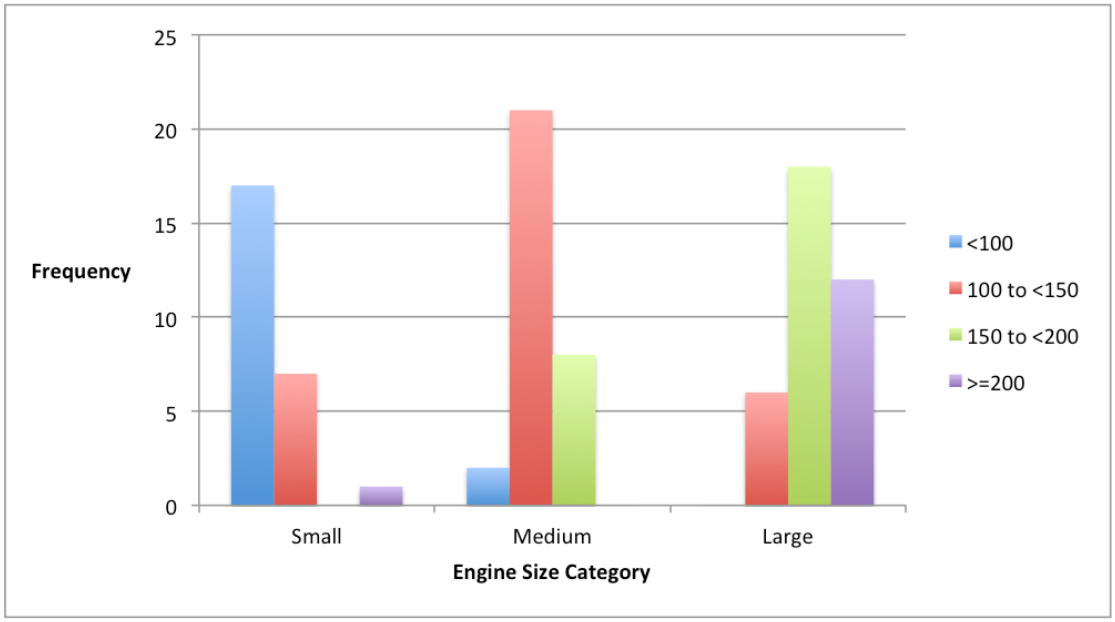


Figure 2.1: Distribution of engine size by horsepower capacity

Using the information we have on the observed frequencies, we compute the expected frequencies of engine size by horsepower category.

Expected Frequencies in a Chi-Square Test for Independence

$$Expected\ Frequency = \frac{(row\ total)(column\ total)}{table\ total}$$

Determining the Expected Count in a Test for Independence

To conduct this test, we will need to compute Expected Counts (E_i) and compare that to Observed Counts (O_i).

Engine Size Category (Oi)	<100	100 to <150	150 to <200	>=200	Total	Relative Frequency
Small	17	7	0	1	25	0.27173913
Medium	2	21	8	0	31	0.336956522
Large	0	6	18	12	36	0.391304348
Total	19	34	26	13	92	
Relative Frequency	0.206521739	0.369565217	0.282608696	0.141304348		1

Table 2.1: Relative marginal frequencies for engine size and horsepower categories

Determining the Expected Proportion

Expected Proportion	<100	100 to <150	150 to <200	>=200
Small	0.056120038	0.100425331	0.076795841	0.038397921
Medium	0.069588847	0.12452741	0.095226843	0.047613422
Large	0.080812854	0.144612476	0.110586011	0.055293006

Table 2.2: Summary table for expected proportion

Determining the Expected Count (Expected Frequencies)

Expected Count (Ei)	<100	100 to <150	150 to <200	>=200
Small	5.163043478	9.239130435	7.065217391	3.532608696
Medium	6.402173913	11.45652174	8.760869565	4.380434783
Large	7.434782609	13.30434783	10.17391304	5.086956522

Table 2.3: Summary table of expected count

From Table 2.3, we are able to verify that all expected frequencies are greater than 1 and that only 1 out of 12 (< 20%) of the expected frequencies has count less than 5, meeting the second test requirement.

Calculating the Test Statistic

Now that we have verified that all the test requirements have been met, we proceed to test the hypotheses regarding the association between horsepower and engine size. We first determine the null and alternative hypothesis and select a level of significance, α.

For the purposes of this report, we determine the null and alternative hypotheses as below:

H_0: Horsepower and Engine Size are independent
H_1: Horsepower and Engine Size are dependent.

We select the Significance Level, α, as 0.05.

Using the **Classical Approach**, we compute the test statistic using the below formula

$$X_0^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$
$$= 36.56134318 + 15.42335733 + 26.8597285$$
$$= 78.84442901$$

Degree of freedom = (number of rows-1)(number of columns -1) = (3-1)(4-1) = 6

Using the Chi-Square Distribution table in Appendix III, the chi-square value for 0.05 with 6 degrees of freedom is 12.592.

This is a summary of Statistic in Classical Approach

χ2 Statistic	78.84442901
Level of Significance	95%
DOF	6
Critical Value	12.592

Since our test-statistic 78.844 > than critical value 12.592, we **reject the null hypothesis**.

Using the **P-value Approach**, we compute the test statistic using the below formula

$$X_0^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$
$$= 36.56134318 + 15.42335733 + 26.8597285$$
$$= 78.84442901$$

Using the Chi-Square Distribution table in Appendix III, we find the row that corresponds to 6 degrees of freedom. The value of 78.844 is greater than 18.548. The area under the chi-square distribution to the right of 18.548 is 0.005. Because 78.844 is greater than 18.548, the P-value is less than 0.005, so we have P-value < 0.005. Alternatively, using R, we are able to determine the p-value of 6.189e-15.

Because the P-value is smaller than the level of significance (0.0005 < 0.05), **we reject the null hypothesis**.

Test Conclusion and Interpretation Statement

The sample data suggest that there is sufficient evidence at the α = 0.05 level of significance to conclude that Horsepower and Engine Size are dependent. We conclude that horsepower and engine size are associated with each other.

Summary Table	
Null hypothesis	Horsepower and Engine Size are independent
Alternative hypothesis	Horsepower and Engine Size are dependent.
Level of significance	95%

Test Statistic	78.844
Critical Value	12.592
Compare critical value w/ test statistic	78.844 > 12.592
Test conclusion	Reject H_0
Interpretation Statement	Since our test-statistic 78.844 > than critical value 12.592, there is sufficient evidence at a 95% level of significance to support the statement that Horsepower and Engine Size are dependent.

Table 2.4: Summary of Q2 analysis

3. Is there a relationship between car horsepower and engine size?

Now that we have established that there is a significant difference between the horsepower in cars with medium and large engines, as well as verified that there is an association between horsepower and engine size, we would like to examine the relationship between car horsepower and engine size.

Table 3.1 shows the list of engine size and horsepower for the upcoming year's car models based on the cars sample data provided in Appendix I.

Engine Size	Horsepower
1.8 1 1.5 1.5 1.8 1.6 1.8 1.5 1.6 1.2 1.8 1.3 1.5 1.8 1.6 1.5 1.3 1.5 1.5 1.3 1.8 1.6 1.9 1.6 1.8 2.8 2.2 2.5 2.3 2.4 2.1 2.2 2 2.3 2.8 2 2.2 2.4 2.3 2.2 2.8 2.4 2.5 2.2 2.2 2.5 2.3 2.3 2 2.2 2.2 2.5 2.2 2.2 2 2.3 3.2 3.5 4.5 3 3 3 3 3 3 3 3.8 5.7 4.9 5 3.3 3.3 3.5 4.6 4.6 3.8 3.8 3.8 4.6 3 3.8 3.8 3.4 3.4 5.7 3 3.4 3.8 4.3 3 3 3.8	140 55 102 81 124 82 103 92 110 73 90 70 82 81 90 92 255 92 92 63 127 74 85 100 92 172 140 164 130 150 140 130 134 114 172 128 130 168 160 135 178 138 109 110 110 100 96 155 110 110 110 100 110 93 115 105 200 208 278 185 225 202 160 155 151 141 170 180 200 170 153 147 214 190 210 170 170 170 295 140 160 140 200 160 300 300 160 170 165 142 145 170

Table 3.1: Engine size and horsepower data

Hypothesis Test Regarding the Slope Coefficient, β_1

Our final analysis will focus on testing whether a linear relationship exists between car horsepower and engine size. Specifically, we will employ a hypothesis test regarding the Slope Coefficient (β_1) to test if a linear relationship exists between the two variables.

Test Requirements

- The sample is obtained using random sampling or from a randomized experiment
- The residuals are normally distributed with constant error variance

Verifying the Data Distribution

Before we begin our test, we need to verify that the residuals are normally distributed with constant error variance by preparing a normal probability plot. This however, requires a few additional steps, as we would need to first identify the least-squares regression equation and the standard error of the estimates.

Finding the least-squares regression equation

We use Microsoft Excel to generate a regression plot based on the data of engine size and horsepower as shown in Figure 3.1. A complete table of the calculated values can be found in the Appendix section under Appendix IX: Summary of residual calculation between horsepower and engine size.

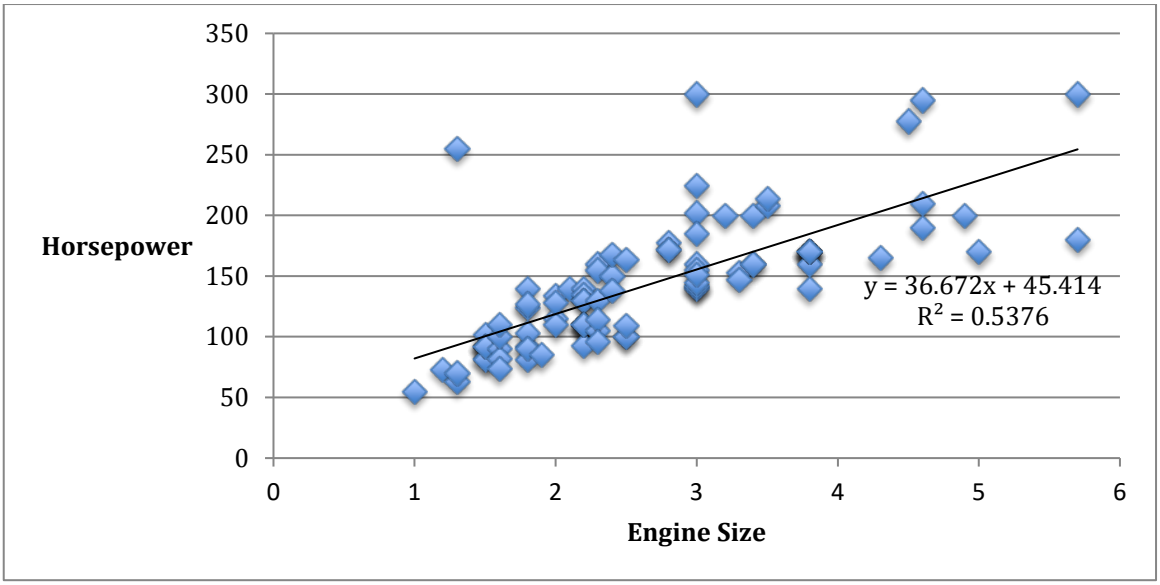


Figure 3.1: Least-squares regression equation on scatter diagram of horsepower and engine size

From Figure 3.1, we find that the least-squares regression equation that best represents the data is given as **y = 36.672x + 45.414.** This equation allows us to deduce the following:

- The regression line minimizes the sum of square differences between the observed values and the predicted values
- The regression coefficient 36.372 is the average change in the response variable (horsepower) per unit change in the explanatory variable (engine size).
- Y-intercept tells us that the minimum horsepower (when engine size is 0) is 45.414.

R and R-squared Discussion

From Figure 3.1, we have also calculated the coefficient of correlation (*R*), as well as the coefficient of determination (R^2).

***R* = 0.733** ***R*² = 0.538**

These help us deduce the following:

- A correlation coefficient 0.733 indicates a strong positive linear relationship between the explanatory variable (engine size) and the response variable (horsepower).
- A coefficient of determination value of 0.538 indicates that 53.8% of the response variables (horsepower) can be predicted by the explanatory variable (engine size).

Now that we have our residuals data ready, we construct a normal probability plot to assess normality.

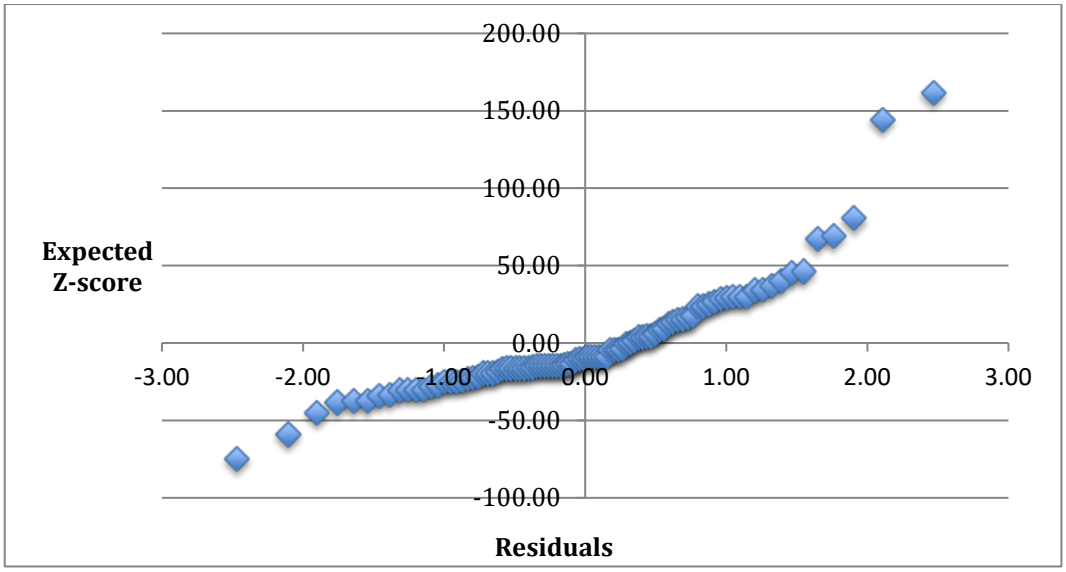


Figure 3.2: Normal probability plot of residuals

For this test, the sample was obtained using random sampling. The normal probability plot of residuals in Figure 3.2 has a correlation of 0.913, suggesting that it is reasonable to conclude that the residuals are normally distributed. Further, these procedures are robust, and minor departures from normality will not adversely affect the results. Since we have a large sample of 92 (more than 30), inferential procedures regarding β_1 can be used even with significant departures from normality. Therefore, we can perform inference on the least-squares regression equation.

Calculating the Test Statistic

Now that we have all the requirements for the test met, we proceed to test whether the variables, horsepower and engine sizes, are linearly related by determining the null and alternative hypotheses and selecting a level of significance.

Two-Tailed	Left-Tailed	Right-Tailed
$H_0: \beta_1 = 0$	$H_0: \beta_1 = 0$	$H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$	$H_1: \beta_1 < 0$	$H_1: \beta_1 > 0$

Figure 3.3: Test hypotheses regarding a slope coefficient

For the purposes of this report, we select a two-tailed hypothesis test and will use a Significance Level, α, of 0.05.

Summary Table	
Null hypothesis	No linear relationship exists between horsepower and engine size, H_0: $\boldsymbol{\beta}_1 = 0$
Alternative hypothesis	A linear relationship exists between horsepower and engine size, H_1: $\boldsymbol{\beta}_1 \neq 0$
Line of regression	y = 36.672x + 45.414
Coefficient of correlation	0.733
Coefficient of determination	0.538
Level of significance	95%
Degrees of freedom	90

Table 3.2: Summary table of data analysis for Q3

Classical Approach

Computing the Standard Error of the Estimates

To test whether a linear relationship exists between car horsepower and engine size at the α = 0.05 level of significance, we need to compute the standard error of the estimate (Se) using the following formula:

$$\text{Standard Error of Estimates}, Se = \sqrt{\frac{\sum residuals^2}{n-2}}$$

$$= \sqrt{\frac{114195.1811}{90}}$$

$$= 35.6207$$

Computing the Sampling Distribution

To test this hypothesis, we also need to know the sampling distribution, S_{b1}. S_{b1} can be calculated using the following formula:

$$Sampling\ distribution, s_{b1} = \frac{s_e}{\sqrt{\sum(x-\bar{x})^2}}$$

$$= \frac{35.6207}{\sqrt{98.7168}}$$

$$= 3.5851$$

Using these, we are able to determine the test statistic using the following formula:

$$t_0 = \frac{b_1 - \beta_1}{s_{b1}} = \frac{b_1}{s_{b1}}$$

which follows Student's *t*-distribution with *n*-2 degrees of freedom.

From the data in Appendix IX, we determine the test statistic (t_0) to be **10.145.**

Because this is a two-tailed test, we determine the critical t-values at the α = 0.05 level of significance with n – 2 = 92 – 2 = 90 degrees of freedom to be $-t_{0.025}$ = -1.987 and $t_{0.025}$ = 1.987.

Since the test statistic, $t_0 = 10.145$ is greater than the critical value $t_{0.025} = 1.987$, we reject the null hypothesis.

P-Value Approach

We have determined from the Classical Approach that the test statistic, $t_0 = 10.145$. Because this a two-tailed test, the P-value is the sum of the area under the *t*-distribution with 90 degrees of freedom to the left of $-t_0 = -10.145$, and to the right of $t_0=10.145$. From the *t*-distribution table, we find that the P value is less than 2(0.0005), which is less than 0.001.

The P-value suggests that about less than 1 sample in 1000 would yield a slope estimate that is as extreme as or more extreme than the one obtained if the null hypothesis of no linear relation were true. Because the P-value is less than 0.001, it is also less than the level of significance $\alpha = 0.05$, and we reject the null hypothesis.

Test Conclusion and Interpretation Statement

There is sufficient evidence at $\alpha = 0.05$ level of significance to conclude that a linear relationship exists between Engine Size and Horsepower.

Summary Table	
Null hypothesis	No linear relationship exists between horsepower and engine size
Alternative hypothesis	A linear relationship exists between horsepower and engine size
Line of regression	y = 36.672x + 45.414
Coefficient of correlation	0.733
Coefficient of determination	0.538
Level of significance	95%
Degrees of freedom	90
Test Statistic	10.145
Critical Values	1.987, - 1.987
Compare critical value w/ test statistic	10.145 > 1.987
P-value	Less than 0.001
Compare P-value to Level of significance	P-value < α = 0.05
Test conclusion	Reject H0

Interpretation Statement	Since Test-statistic 10.145 > critical value1.987, there is sufficient evidence at α = 0.05 level of significance to conclude that a linear relationship exists between Engine Size and Horsepower

Table 3.3: Summary of Q3 analysis

Constructing a (1- α) • 100% Confidence Interval for the Slope of the Regression Line

We will determine with a 95% confidence interval the slope of the true regression line using the following formulas:

$$Lower\ bound = b_1 - t_{\frac{\alpha}{2}} \cdot \frac{s_e}{\sqrt{\sum(x_i - \bar{x})^2}}$$

$$Upper\ bound = b_1 + t_{\frac{\alpha}{2}} \cdot \frac{s_e}{\sqrt{\sum(x_i - \bar{x})^2}}$$

From the data in Appendix IX, at 90 degrees of freedom, we find that the confidence intervals are as follows:

Upper bound: 43.795

Lower bound: 29.549

Interpretation Statement

We are 95% confident that the mean increase in horsepower for each additional unit in engine size is somewhere between 29.549 and 34.795.. Because the 95% confidence interval does not include 0, we reject H_0.

Conclusion

Despite recent advancements in engine and automotive technology, our study suggests that there remains a strong linear correlation between engine size and horsepower, and that the difference in horsepower between cars with medium sized engines and cars with large sized engines are significant.

In the future, we would like to repeat the statistical test on correlation between engine size and horsepower with new data sets to determine if a linear relation continues to exist between these variables. If the advancement in technology in the automotive industry is reflected in the new data set, we may hypothesize that new car data will reflect higher horsepower with smaller engine sizes, thereby weakening the linear relationship we found in this study.

To better understand other variables that affect horsepower, we would suggest future research into how the engine size and car weight affect horsepower. For this, we would suggest using a multiple regression model.

We can also extend our research to find out correlations between price and other variables. To be specific, we can conduct research to find out how price relates to multiple variables such as engine size, horsepower and MPG, via a multiple regression model. By comparing these results, we can best identify how each variable affects price and use this to define pricing strategies in the automotive industry.

Lessons Learned

- Different people using different software can come up with slightly different results, though these are usually not significant enough to affect test conclusions.

- People read and interpret data differently – if we did not agree upon a fixed methodology upfront, we realized that different people would interpret data in different ways (e.g. For question 3, some people used horsepower as the dependent variable and engine size as the independent variable, whilst others used horsepower as the independent variable and engine size as the dependent variable).

- The group project was a great way for us to familiarize ourselves with the entire process of statistical analysis, including selection of topics, method selection, numerical analysis, and derivation of conclusions.

- In this group project, we only used a simple regression model., If we are able to use multiple regression analysis approaches in the future, it will allow us to explicitly control for many other variables and allow us to make more complex predictions for our dependent variables.

Reference

Deloitte (2014, December). Driving through the consumer's mind: Steps in the buying process Retrieved from https://www2.deloitte.com/content/dam/Deloitte/in/Documents/manufacturing/in-mfg-dtcm-steps-in-the-buying-process-noexp.pdf

Deloitte (2014, February). Driving through the consumer's mind: Considerations for Car purchase Retrieved from https://www2.deloitte.com/content/dam/Deloitte/in/Documents/manufacturing/deloitte-india-mfg-driving-through-consumers-mind-noexp.pdf

Appendix

Appendix I: Original cars dataset

Manufacturer	Origin	Model	Type	Passengers	Price	Price Category	CityMPG	City MPG Category	HighwayMPG	HighwayMPG Category	Engine Size	Engine Size Category	Horsepower	Horsepower Category	Fuel Tank	Weight	Weight Category
Audi	non-US	90	Compact	5	29.1	Expensive	20	Poor	26	Poor	2.8	Medium	172	150 to <200	16.9	3375	> 1.5 Ton
Honda	non-US	Accord	Compact	4	17.5	Moderate	24	Good	31	Good	2.2	Medium	140	100 to <150	17	3040	< 1.5 Ton
Mazda	non-US	626	Compact	5	16.5	Moderate	26	Good	34	Good	2.5	Medium	164	150 to <200	15.5	2970	< 1.5 Ton
Mercedes-Benz	non-US	190E	Compact	5	31.9	Expensive	20	Poor	29	Poor	2.3	Medium	130	100 to <150	14.5	2920	< 1.5 Ton
Nissan	non-US	Altima	Compact	5	15.7	Moderate	24	Good	30	Good	2.4	Medium	150	150 to <200	15.9	3050	> 1.5 Ton
Saab	non-US	900	Compact	5	28.7	Expensive	20	Poor	26	Poor	2.1	Medium	140	100 to <150	18	2775	< 1.5 Ton
Subaru	non-US	Legacy	Compact	5	19.5	Moderate	23	Good	30	Good	2.2	Medium	130	100 to <150	15.9	3085	> 1.5 Ton
Volkswagen	non-US	Passat	Compact	5	20	Expensive	21	Good	30	Good	2	Medium	134	100 to <150	18.5	2985	< 1.5 Ton
Volvo	non-US	240	Compact	5	22.7	Expensive	21	Good	28	Poor	2.3	Medium	114	100 to <150	15.8	2985	< 1.5 Ton
Acura	non-US	Legend	Midsize	5	33.9	Expensive	18	Poor	25	Poor	3.2	Large	200	>=200	18	3560	> 1.5 Ton
Audi	non-US	100	Midsize	6	37.7	Expensive	19	Poor	26	Poor	2.8	Medium	172	150 to <200	21.1	3405	> 1.5 Ton
BMW	non-US	535i	Midsize	4	30	Expensive	22	Good	30	Good	3.5	Large	208	>=200	21.1	3640	> 1.5 Ton
Hyundai	non-US	Sonata	Midsize	5	13.9	Cheap	20	Poor	27	Poor	2	Medium	128	100 to <150	17.2	2885	< 1.5 Ton
Infiniti	non-US	Q45	Midsize	5	47.9	Expensive	17	Poor	22	Poor	4.5	Large	278	>=200	22.5	4000	> 1.5 Ton
Lexus	non-US	ES300	Midsize	5	28	Expensive	18	Poor	24	Poor	3	Large	185	150 to <200	18.5	3510	> 1.5 Ton
Lexus	non-US	SC300	Midsize	4	35.2	Expensive	18	Poor	23	Poor	3	Large	225	>=200	20.6	3515	> 1.5 Ton
Mitsubishi	non-US	Diamante	Midsize	5	26.1	Expensive	18	Poor	24	Poor	3	Large	202	>=200	19	3730	> 1.5 Ton
Nissan	non-US	Maxima	Midsize	5	21.5	Expensive	21	Good	26	Poor	3	Large	160	150 to <200	18.5	3200	> 1.5 Ton
Toyota	non-US	Camry	Midsize	5	18.2	Moderate	22	Good	29	Poor	2.2	Medium	130	100 to <150	18.5	3030	> 1.5 Ton
Volvo	non-US	850	Midsize	5	25.7	Expensive	20	Poor	28	Poor	2.4	Medium	168	150 to <200	19.3	3245	> 1.5 Ton
Acura	non-US	Integra	Small	5	15.9	Moderate	25	Good	31	Good	1.8	Small	140	100 to <150	13.2	2705	< 1.5 Ton
Geo	non-US	Metro	Small	4	8.4	Cheap	46	Good	50	Good	1	Small	55	<100	10.6	1695	< 1.5 Ton
Honda	non-US	Civic	Small	4	12.1	Cheap	42	Good	46	Good	1.5	Small	102	100 to <150	11.9	2350	< 1.5 Ton
Hyundai	non-US	Excel	Small	5	8	Cheap	29	Good	33	Good	1.5	Small	81	<100	11.9	2345	< 1.5 Ton
Hyundai	non-US	Elantra	Small	5	10	Cheap	22	Good	29	Poor	1.8	Small	124	100 to <150	13.7	2620	< 1.5 Ton
Mazda	non-US	323	Small	4	8.3	Cheap	29	Good	37	Good	1.6	Small	82	<100	13.2	2325	< 1.5 Ton
Mazda	non-US	Protege	Small	5	11.6	Cheap	28	Good	36	Good	1.8	Small	103	100 to <150	14.5	2440	< 1.5 Ton
Mitsubishi	non-US	Mirage	Small	5	10.3	Cheap	29	Good	33	Good	1.5	Small	92	<100	13.2	2295	< 1.5 Ton
Nissan	non-US	Sentra	Small	5	11.8	Cheap	29	Good	33	Good	1.6	Small	110	100 to <150	13.2	2545	< 1.5 Ton
Subaru	non-US	Justy	Small	4	8.4	Cheap	33	Good	37	Good	1.2	Small	73	<100	9.2	2045	< 1.5 Ton
Subaru	non-US	Loyale	Small	5	10.9	Cheap	25	Good	30	Good	1.8	Small	90	<100	15.9	2490	< 1.5 Ton
Suzuki	non-US	Swift	Small	4	8.6	Cheap	39	Good	43	Good	1.3	Small	70	<100	10.6	1965	< 1.5 Ton
Toyota	non-US	Tercel	Small	5	9.8	Cheap	32	Good	37	Good	1.5	Small	82	<100	11.9	2055	< 1.5 Ton
Volkswagen	non-US	Fox	Small	4	9.1	Cheap	25	Good	33	Good	1.8	Small	81	<100	12.4	2240	< 1.5 Ton
Geo	non-US	Storm	Sporty	4	12.5	Cheap	30	Good	36	Good	1.6	Small	90	<100	12.4	2475	< 1.5 Ton
Honda	non-US	Prelude	Sporty	4	19.8	Moderate	24	Good	31	Good	2.3	Medium	160	150 to <200	15.9	2865	< 1.5 Ton
Hyundai	non-US	Scoupe	Sporty	4	10	Cheap	26	Good	34	Good	1.5	Small	92	<100	11.9	2285	< 1.5 Ton
Mazda	non-US	RX-7	Sporty	2	32.5	Expensive	17	Poor	25	Poor	1.3	Small	255	>=200	20	2895	< 1.5 Ton
Toyota	non-US	Celica	Sporty	4	18.4	Moderate	25	Good	32	Good	2.2	Medium	135	100 to <150	15.9	2950	< 1.5 Ton
Volkswagen	non-US	Corrado	Sporty	4	23.3	Expensive	18	Poor	25	Poor	2.8	Medium	178	150 to <200	18.5	2810	< 1.5 Ton
Mazda	non-US	MPV	Van	7	19.1	Moderate	18	Poor	24	Poor	3	Large	155	150 to <200	19.6	3735	> 1.5 Ton
Nissan	non-US	Quest	Van	7	19.1	Moderate	17	Poor	23	Poor	3	Large	151	150 to <200	20	4100	> 1.5 Ton
Toyota	non-US	Previa	Van	7	22.7	Expensive	18	Poor	22	Poor	2.4	Medium	138	100 to <150	19.8	3785	> 1.5 Ton
Volkswagen	non-US	Eurovan	Van	7	19.7	Moderate	17	Poor	21	Poor	2.5	Medium	109	100 to <150	21.1	3960	> 1.5 Ton
Chevrolet	US	Cavalier	Compact	5	13.4	Cheap	25	Good	36	Good	2.2	Medium	110	100 to <150	15.2	2490	< 1.5 Ton
Chevrolet	US	Corsica	Compact	5	11.4	Cheap	25	Good	34	Good	2.2	Medium	110	100 to <150	15.6	2785	< 1.5 Ton
Chrysler	US	LeBaron	Compact	6	15.8	Moderate	23	Good	28	Poor	3	Large	141	100 to <150	16	3085	> 1.5 Ton
Dodge	US	Spirit	Compact	6	13.3	Cheap	22	Good	27	Poor	2.5	Medium	100	100 to <150	16	2970	< 1.5 Ton
Ford	US	Tempo	Compact	5	11.3	Cheap	22	Good	27	Poor	2.3	Medium	96	<100	15.9	2690	< 1.5 Ton
Oldsmobile	US	Achieva	Compact	5	13.5	Cheap	24	Good	31	Good	2.3	Medium	155	150 to <200	15.2	2910	< 1.5 Ton
Pontiac	US	Sunbird	Compact	5	11.1	Cheap	23	Good	31	Good	2	Medium	110	100 to <150	15.2	2575	< 1.5 Ton
Buick	US	LeSabre	Large	6	20.8	Expensive	19	Poor	28	Poor	3.8	Large	170	150 to <200	18	3470	> 1.5 Ton
Buick	US	Roadmaster	Large	6	23.7	Expensive	16	Poor	25	Poor	5.7	Large	180	150 to <200	23	4105	> 1.5 Ton
Cadillac	US	DeVille	Large	6	34.7	Expensive	16	Poor	25	Poor	4.9	Large	200	>=200	18	3620	> 1.5 Ton
Chevrolet	US	Caprice	Large	6	18.8	Moderate	17	Poor	26	Poor	5	Large	170	150 to <200	23	3910	> 1.5 Ton
Chrysler	US	Concorde	Large	6	18.4	Moderate	20	Poor	28	Poor	3.3	Large	153	150 to <200	18	3515	> 1.5 Ton
Chrysler	US	Imperial	Large	6	29.5	Expensive	20	Poor	26	Poor	3.3	Large	147	100 to <150	16	3570	> 1.5 Ton
Eagle	US	Vision	Large	6	19.3	Moderate	20	Poor	28	Poor	3.5	Large	214	>=200	18	3490	> 1.5 Ton
Ford	US	Crown_Vict	Large	6	20.9	Expensive	18	Poor	26	Poor	4.6	Large	190	150 to <200	20	3950	> 1.5 Ton
Lincoln	US	Town_Car	Large	6	38.1	Expensive	18	Poor	26	Poor	4.6	Large	210	>=200	20	4055	> 1.5 Ton
Oldsmobile	US	Eighty-Eig	Large	6	20.7	Expensive	19	Poor	28	Poor	3.8	Large	170	150 to <200	18	3470	> 1.5 Ton
Pontiac	US	Bonneville	Large	6	24.4	Expensive	19	Poor	28	Poor	3.8	Large	170	150 to <200	18	3495	> 1.5 Ton
Buick	US	Century	Midsize	6	15.7	Moderate	22	Good	31	Good	2.2	Medium	110	100 to <150	16.4	2880	< 1.5 Ton
Buick	US	Riviera	Midsize	5	28.3	Expensive	19	Poor	27	Poor	3.8	Large	170	150 to <200	18.8	3495	> 1.5 Ton
Cadillac	US	Seville	Midsize	5	40.1	Expensive	16	Poor	25	Poor	4.6	Large	295	>=200	20	3935	> 1.5 Ton
Chevrolet	US	Lumina	Midsize	6	15.9	Moderate	21	Good	29	Poor	2.2	Medium	110	100 to <150	16.5	3195	> 1.5 Ton
Dodge	US	Dynasty	Midsize	6	15.6	Moderate	21	Good	27	Poor	2.5	Medium	100	100 to <150	16	3080	> 1.5 Ton
Ford	US	Taurus	Midsize	5	20.2	Expensive	21	Good	30	Good	3	Large	140	100 to <150	16	3325	> 1.5 Ton
Lincoln	US	Continenta	Midsize	6	34.3	Expensive	17	Poor	26	Poor	3.8	Large	160	150 to <200	18.4	3695	> 1.5 Ton
Mercury	US	Cougar	Midsize	5	14.9	Cheap	19	Poor	26	Poor	3.8	Large	140	100 to <150	18	3610	> 1.5 Ton
Oldsmobile	US	Cutlass_Ci	Midsize	5	16.3	Moderate	23	Good	31	Good	2.2	Medium	110	100 to <150	16.5	2890	< 1.5 Ton
Pontiac	US	Grand_Prix	Midsize	5	18.5	Moderate	19	Poor	27	Poor	3.4	Large	200	>=200	16.5	3450	> 1.5 Ton
Dodge	US	Colt	Small	5	9.2	Cheap	29	Good	33	Good	1.5	Small	92	<100	13.2	2270	< 1.5 Ton
Dodge	US	Shadow	Small	5	11.3	Cheap	23	Good	29	Poor	2.2	Medium	93	<100	14	2670	< 1.5 Ton
Eagle	US	Summit	Small	5	12.2	Cheap	29	Good	33	Good	1.5	Small	92	<100	13.2	2295	< 1.5 Ton
Ford	US	Festiva	Small	4	7.4	Cheap	31	Good	33	Good	1.3	Small	63	<100	10	1845	< 1.5 Ton
Ford	US	Escort	Small	5	10.1	Cheap	23	Good	30	Good	1.8	Small	127	100 to <150	13.2	2530	< 1.5 Ton
Pontiac	US	LeMans	Small	4	9	Cheap	31	Good	41	Good	1.6	Small	74	<100	13.2	2350	< 1.5 Ton
Saturn	US	SL	Small	5	11.1	Cheap	28	Good	38	Good	1.9	Small	85	<100	12.8	2495	< 1.5 Ton
Chevrolet	US	Camaro	Sporty	4	15.1	Moderate	19	Poor	28	Poor	3.4	Large	160	150 to <200	15.5	3240	> 1.5 Ton
Chevrolet	US	Corvette	Sporty	2	38	Expensive	17	Poor	25	Poor	5.7	Large	300	>=200	20	3380	> 1.5 Ton
Dodge	US	Stealth	Sporty	4	25.8	Expensive	18	Poor	24	Poor	3	Large	300	>=200	19.8	3805	> 1.5 Ton
Ford	US	Probe	Sporty	4	14	Cheap	24	Good	30	Good	2	Medium	115	100 to <150	15.5	2710	< 1.5 Ton
Ford	US	Mustang	Sporty	4	15.9	Moderate	22	Good	29	Poor	2.3	Medium	105	100 to <150	15.4	2850	< 1.5 Ton
Mercury	US	Capri	Sporty	4	14.1	Cheap	23	Good	26	Poor	1.6	Small	100	100 to <150	11.1	2450	< 1.5 Ton
Plymouth	US	Laser	Sporty	4	14.4	Cheap	23	Good	30	Good	1.8	Small	92	<100	15.9	2640	< 1.5 Ton
Pontiac	US	Firebird	Sporty	4	17.7	Moderate	19	Poor	28	Poor	3.4	Large	160	150 to <200	15.5	3240	> 1.5 Ton
Chevrolet	US	Lumina_APV	Van	7	16.3	Moderate	18	Poor	23	Poor	3.8	Large	170	150 to <200	20	3715	> 1.5 Ton
Chevrolet	US	Astro	Van	8	16.6	Moderate	15	Poor	20	Poor	4.3	Large	165	150 to <200	27	4025	> 1.5 Ton
Dodge	US	Caravan	Van	7	19	Moderate	17	Poor	21	Poor	3	Large	142	100 to <150	20	3705	> 1.5 Ton
Ford	US	Aerostar	Van	7	19.9	Moderate	15	Poor	20	Poor	3	Large	145	100 to <150	21	3735	> 1.5 Ton
Oldsmobile	US	Silhouette	Van	7	19.5	Moderate	18	Poor	23	Poor	3.8	Large	170	150 to <200	20	3715	> 1.5 Ton

Appendix II: Critical Values for Student's *t*-Distributions

TABLE of CRITICAL VALUES for STUDENT'S *t* DISTRIBUTIONS

Column headings denote probabilities (α) ***above*** tabulated values.

d.f.	0.40	0.25	0.10	0.05	0.04	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	7.916	12.706	15.894	31.821	63.656	127.321	318.289	636.578
2	0.289	0.816	1.886	2.920	3.320	4.303	4.849	6.965	9.925	14.089	22.328	31.600
3	0.277	0.765	1.638	2.353	2.605	3.182	3.482	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.333	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.191	2.571	2.757	3.365	4.032	4.773	5.894	6.869
6	0.265	0.718	1.440	1.943	2.104	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.046	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.004	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	1.973	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	1.948	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	1.928	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	1.912	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	1.899	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	1.887	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	1.878	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	1.869	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	1.862	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	1.855	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	1.850	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	1.844	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	1.840	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	1.835	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	1.832	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.256	0.685	1.318	1.711	1.828	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	1.825	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	1.822	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	1.819	2.052	2.158	2.473	2.771	3.057	3.421	3.689
28	0.256	0.683	1.313	1.701	1.817	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	1.814	2.045	2.150	2.462	2.756	3.038	3.396	3.660
30	0.256	0.683	1.310	1.697	1.812	2.042	2.147	2.457	2.750	3.030	3.385	3.646
31	0.256	0.682	1.309	1.696	1.810	2.040	2.144	2.453	2.744	3.022	3.375	3.633
32	0.255	0.682	1.309	1.694	1.808	2.037	2.141	2.449	2.738	3.015	3.365	3.622
33	0.255	0.682	1.308	1.692	1.806	2.035	2.138	2.445	2.733	3.008	3.356	3.611
34	0.255	0.682	1.307	1.691	1.805	2.032	2.136	2.441	2.728	3.002	3.348	3.601
35	0.255	0.682	1.306	1.690	1.803	2.030	2.133	2.438	2.724	2.996	3.340	3.591
36	0.255	0.681	1.306	1.688	1.802	2.028	2.131	2.434	2.719	2.990	3.333	3.582
37	0.255	0.681	1.305	1.687	1.800	2.026	2.129	2.431	2.715	2.985	3.326	3.574
38	0.255	0.681	1.304	1.686	1.799	2.024	2.127	2.429	2.712	2.980	3.319	3.566
39	0.255	0.681	1.304	1.685	1.798	2.023	2.125	2.426	2.708	2.976	3.313	3.558
40	0.255	0.681	1.303	1.684	1.796	2.021	2.123	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	1.781	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.254	0.678	1.292	1.664	1.773	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.254	0.677	1.290	1.660	1.769	1.984	2.081	2.364	2.626	2.871	3.174	3.390
120	0.254	0.677	1.289	1.658	1.766	1.980	2.076	2.358	2.617	2.860	3.160	3.373
140	0.254	0.676	1.288	1.656	1.763	1.977	2.073	2.353	2.611	2.852	3.149	3.361
160	0.254	0.676	1.287	1.654	1.762	1.975	2.071	2.350	2.607	2.847	3.142	3.352
180	0.254	0.676	1.286	1.653	1.761	1.973	2.069	2.347	2.603	2.842	3.136	3.345
200	0.254	0.676	1.286	1.653	1.760	1.972	2.067	2.345	2.601	2.838	3.131	3.340
250	0.254	0.675	1.285	1.651	1.758	1.969	2.065	2.341	2.596	2.832	3.123	3.330
inf	0.253	0.674	1.282	1.645	1.751	1.960	2.054	2.326	2.576	2.807	3.090	3.290

Appendix III: Chi-Square Distribution

Chi-Square (χ^2) Distribution

Degrees of Freedom	Area to the Right of Critical Value							
	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892

Appendix IV: Normal probability data for horsepower of cars with medium engines

Horsepower	Number	Cumulative Probability	z value
93	1	0.016129032	-2.141198121
96	2	0.048387097	-1.660697611
100	3	0.080645161	-1.400745061
100	4	0.112903226	-1.211232131
105	5	0.14516129	-1.057414228
109	6	0.177419355	-0.92524456
110	7	0.209677419	-0.807541042
110	8	0.241935484	-0.700090213
110	9	0.274193548	-0.600178776
110	10	0.306451613	-0.505933654
110	11	0.338709677	-0.41598722
110	12	0.370967742	-0.329291347
114	13	0.403225806	-0.245006223
115	14	0.435483871	-0.162429373
128	15	0.467741935	-0.08094729
130	16	0.5	0
130	17	0.532258065	0.08094729
130	18	0.564516129	0.162429373
134	19	0.596774194	0.245006223
135	20	0.629032258	0.329291347
138	21	0.661290323	0.41598722
140	22	0.693548387	0.505933654
140	23	0.725806452	0.600178776
150	24	0.758064516	0.700090213
155	25	0.790322581	0.807541042
160	26	0.822580645	0.92524456
164	27	0.85483871	1.057414228
168	28	0.887096774	1.211232131
172	29	0.919354839	1.400745061
172	30	0.951612903	1.660697611
178	31	0.983870968	2.141198121
TOTAL	**31**		

Appendix V: Normal probability data for horsepower of cars with large engines

Horsepower	Number	Cumulative Probability	z value
140	1	0.013888889	-2.200410581
140	2	0.041666667	-1.731664396
141	3	0.069444444	-1.479941389
142	4	0.097222222	-1.297542929
145	5	0.125	-1.15034938
147	6	0.152777778	-1.024592393
151	7	0.180555556	-0.913249927
153	8	0.208333333	-0.812217801
155	9	0.236111111	-0.718868053
160	10	0.263888889	-0.631401894
160	11	0.291666667	-0.548522283
160	12	0.319444444	-0.469252884
160	13	0.347222222	-0.392830814
165	14	0.375	-0.318639364
170	15	0.402777778	-0.246163647
170	16	0.430555556	-0.17495994
170	17	0.458333333	-0.104633456
170	18	0.486111111	-0.034821317
170	19	0.513888889	0.034821317
170	20	0.541666667	0.104633456
170	21	0.569444444	0.17495994
180	22	0.597222222	0.246163647
185	23	0.625	0.318639364
190	24	0.652777778	0.392830814
200	25	0.680555556	0.469252884
200	26	0.708333333	0.548522283
200	27	0.736111111	0.631401894
202	28	0.763888889	0.718868053
208	29	0.791666667	0.812217801
210	30	0.819444444	0.913249927
214	31	0.847222222	1.024592393
225	32	0.875	1.15034938
278	33	0.902777778	1.297542929
295	34	0.930555556	1.479941389
300	35	0.958333333	1.731664396
300	36	0.986111111	2.200410581
TOTAL	**36**		

Appendix VI: Summary table of $O_i - E_i$

Oi-Ei	<100	100 to <150	150 to <200	>=200
Small	11.83695652	-2.239130435	-7.065217391	-2.532608696
Medium	-4.402173913	9.543478261	-0.760869565	-4.380434783
Large	-7.434782609	-7.304347826	7.826086957	6.913043478

Appendix VII: Summary table of $(O_i - E_i)^2$

(Oi-Ei)^2	<100	100 to <150	150 to <200	>=200
Small	140.1135397	5.013705104	49.91729679	6.414106805
Medium	19.37913516	91.07797732	0.578922495	19.18820888
Large	55.27599244	53.35349716	61.24763705	47.79017013

Appendix VIII: Summary table of $(O_i - E_i)^2 / E_i$

(Oi-Ei)^2 / Ei	<100	100 to <150	150 to <200	>=200	Total
Small	27.13778032	0.542659847	7.065217391	1.815685619	36.56134318
Medium	3.026961689	7.949880373	0.066080483	4.380434783	15.42335733
Large	7.434782609	4.010230179	6.02006689	9.394648829	26.85972851

Appendix IX: Summary of residual calculation between horsepower and engine size

EngineSize (x)	Horsepower (y)	Y-hat	Residual	Residual^ 2	xi - x-bar	(xi - x-bar)^2
1	55	82.086	-27.086	733.651396	-1.661956522	2.76209948
1.2	73	89.4204	-16.4204	269.6295362	-1.461956522	2.137316871
1.3	255	93.0876	161.9124	26215.62527	-1.361956522	1.854925567
1.3	63	93.0876	-30.0876	905.2636738	-1.361956522	1.854925567
1.3	70	93.0876	-23.0876	533.0372738	-1.361956522	1.854925567
1.5	92	100.422	-8.422	70.930084	-1.161956522	1.350142958
1.5	102	100.422	1.578	2.490084	-1.161956522	1.350142958
1.5	92	100.422	-8.422	70.930084	-1.161956522	1.350142958
1.5	82	100.422	-18.422	339.370084	-1.161956522	1.350142958
1.5	81	100.422	-19.422	377.214084	-1.161956522	1.350142958
1.5	92	100.422	-8.422	70.930084	-1.161956522	1.350142958
1.5	92	100.422	-8.422	70.930084	-1.161956522	1.350142958
1.6	90	104.0892	-14.0892	198.5055566	-1.061956522	1.127751654
1.6	82	104.0892	-22.0892	487.9327566	-1.061956522	1.127751654
1.6	100	104.0892	-4.0892	16.72155664	-1.061956522	1.127751654
1.6	74	104.0892	-30.0892	905.3599566	-1.061956522	1.127751654
1.6	110	104.0892	5.9108	34.93755664	-1.061956522	1.127751654
1.8	81	111.4236	-30.4236	925.595437	-0.861956522	0.742969045
1.8	92	111.4236	-19.4236	377.276237	-0.861956522	0.742969045
1.8	103	111.4236	-8.4236	70.95703696	-0.861956522	0.742969045
1.8	140	111.4236	28.5764	816.610637	-0.861956522	0.742969045
1.8	90	111.4236	-21.4236	458.970637	-0.861956522	0.742969045
1.8	124	111.4236	12.5764	158.165837	-0.861956522	0.742969045
1.8	127	111.4236	15.5764	242.624237	-0.861956522	0.742969045
1.9	85	115.0908	-30.0908	905.4562446	-0.761956522	0.580577741
2	115	118.758	-3.758	14.122564	-0.661956522	0.438186437
2	134	118.758	15.242	232.318564	-0.661956522	0.438186437
2	110	118.758	-8.758	76.702564	-0.661956522	0.438186437
2	128	118.758	9.242	85.414564	-0.661956522	0.438186437
2.1	140	122.4252	17.5748	308.873595	-0.561956522	0.315795132
2.2	140	126.0924	13.9076	193.4213378	-0.461956522	0.213403828
2.2	135	126.0924	8.9076	79.34533776	-0.461956522	0.213403828
2.2	110	126.0924	-16.0924	258.9653378	-0.461956522	0.213403828
2.2	130	126.0924	3.9076	15.26933776	-0.461956522	0.213403828
2.2	130	126.0924	3.9076	15.26933776	-0.461956522	0.213403828
2.2	110	126.0924	-16.0924	258.9653378	-0.461956522	0.213403828
2.2	93	126.0924	-33.0924	1095.106938	-0.461956522	0.213403828
2.2	110	126.0924	-16.0924	258.9653378	-0.461956522	0.213403828
2.2	110	126.0924	-16.0924	258.9653378	-0.461956522	0.213403828
2.2	110	126.0924	-16.0924	258.9653378	-0.461956522	0.213403828
2.3	160	129.7596	30.2404	914.4817922	-0.361956522	0.131012524
2.3	105	129.7596	-24.7596	613.0377922	-0.361956522	0.131012524
2.3	155	129.7596	25.2404	637.0777922	-0.361956522	0.131012524
2.3	130	129.7596	0.2404	0.05779216	-0.361956522	0.131012524
2.3	114	129.7596	-15.7596	248.3649922	-0.361956522	0.131012524
2.3	96	129.7596	-33.7596	1139.710592	-0.361956522	0.131012524
2.4	150	133.4268	16.5732	274.6709582	-0.261956522	0.068621219
2.4	168	133.4268	34.5732	1195.306158	-0.261956522	0.068621219
2.4	138	133.4268	4.5732	20.91415824	-0.261956522	0.068621219
2.5	164	137.094	26.906	723.932836	-0.161956522	0.026229915
2.5	100	137.094	-37.094	1375.964836	-0.161956522	0.026229915
2.5	100	137.094	-37.094	1375.964836	-0.161956522	0.026229915
2.5	109	137.094	-28.094	789.272836	-0.161956522	0.026229915

2.8	178	148.0956	29.9044	894.2731394	0.138043478	0.019056002
2.8	172	148.0956	23.9044	571.4203394	0.138043478	0.019056002
2.8	172	148.0956	23.9044	571.4203394	0.138043478	0.019056002
3	300	155.43	144.57	20900.4849	0.338043478	0.114273393
3	225	155.43	69.57	4839.9849	0.338043478	0.114273393
3	160	155.43	4.57	20.8849	0.338043478	0.114273393
3	202	155.43	46.57	2168.7649	0.338043478	0.114273393
3	140	155.43	-15.43	238.0849	0.338043478	0.114273393
3	185	155.43	29.57	874.3849	0.338043478	0.114273393
3	141	155.43	-14.43	208.2249	0.338043478	0.114273393
3	142	155.43	-13.43	180.3649	0.338043478	0.114273393
3	155	155.43	-0.43	0.1849	0.338043478	0.114273393
3	145	155.43	-10.43	108.7849	0.338043478	0.114273393
3	151	155.43	-4.43	19.6249	0.338043478	0.114273393
3.2	200	162.7644	37.2356	1386.489907	0.538043478	0.289490784
3.3	153	166.4316	-13.4316	180.4078786	0.638043478	0.40709948
3.3	147	166.4316	-19.4316	377.5870786	0.638043478	0.40709948
3.4	160	170.0988	-10.0988	101.9857614	0.738043478	0.544708176
3.4	160	170.0988	-10.0988	101.9857614	0.738043478	0.544708176
3.4	200	170.0988	29.9012	894.0817614	0.738043478	0.544708176
3.5	208	173.766	34.234	1171.966756	0.838043478	0.702316871
3.5	214	173.766	40.234	1618.774756	0.838043478	0.702316871
3.8	140	184.7676	-44.7676	2004.13801	1.138043478	1.295142958
3.8	170	184.7676	-14.7676	218.0820098	1.138043478	1.295142958
3.8	170	184.7676	-14.7676	218.0820098	1.138043478	1.295142958
3.8	160	184.7676	-24.7676	613.4340098	1.138043478	1.295142958
3.8	170	184.7676	-14.7676	218.0820098	1.138043478	1.295142958
3.8	170	184.7676	-14.7676	218.0820098	1.138043478	1.295142958
3.8	170	184.7676	-14.7676	218.0820098	1.138043478	1.295142958
3.8	170	184.7676	-14.7676	218.0820098	1.138043478	1.295142958
4.3	165	203.1036	-38.1036	1451.884333	1.638043478	2.683186437
4.5	278	210.438	67.562	4564.623844	1.838043478	3.378403828
4.6	295	214.1052	80.8948	6543.968667	1.938043478	3.756012524
4.6	190	214.1052	-24.1052	581.060667	1.938043478	3.756012524
4.6	210	214.1052	-4.1052	16.85266704	1.938043478	3.756012524
4.9	200	225.1068	-25.1068	630.3514062	2.238043478	5.008838611
5	170	228.774	-58.774	3454.383076	2.338043478	5.466447306
5.7	300	254.4444	45.5556	2075.312691	3.038043478	9.229708176
5.7	180	254.4444	-74.4444	5541.968691	3.038043478	9.229708176

Sum of Residual^2	Divided by N-2
114195.1811	1268.835346

X-bar	Summation (xi - x-bar)^2	Sqrt
2.661956522	98.71684783	9.93563525

Se	Sb1	b1
35.62071512	3.585147223	36.372

Test-Statistic
10.14519007

Appendix X: R code snippet for Q1

```
large_horsepower<-read.csv("LARGE_HORSEPOWER.csv")
head(large_horsepower)
summary(large_horsepower)
medium_horsepower<-read.csv("MEDIUM_HORSEPOWER.csv")
head(medium_horsepower)
summary(medium_horsepower)
plot(large_horsepower$Enginesize_Category, large_horsepower$Horsepower)

plot(medium_horsepower$Enginesize_Category, medium_horsepower$Horsepower)
```

Appendix XI: R code snippet for Q2

```
> tbl = table(cars$Enginesize_Category,
cars$Horsepower_Category)

# preparing the data in a table
> tbl:
        <100 >=200 100 to <150 150 to <200
  Large     0    12          6          18
  Medium    2     0         21           8
  Small    17     1          7           0

# computing the chi-square test
> chisq.test(tbl)
X-squared = 78.844, df = 6, p-value = 6.189e-
15
```

Appendix XII: R code snippet for Q3

```
# loading the dataset
horsepowerenginesize<-read.csv("horsepower_enginesize_Rversion.csv")
head(horsepowerenginesize)
summary(horsepowerenginesize)

# scatterplot
scatter.smooth(x=horsepowerenginesize$Horsepower, y=horsepowerenginesize$EngineSize,
main="Horsepower ~ Enginesize")
result=lm(EngineSize~Horsepower,data=horsepowerenginesize)
summary(result)
```