DSCC/CSC/TCS 462 Statistics Assignment 1

Aradhya Mathur

2022-09-19

```
library(readr)
library(ggplot2)
library(moments)
```

Question 1) For the first part of this assignment, we will explore the relationships between variables using the same "car_sales.csv" dataset as HW0. In particular, we will explore the relationships between multiple variables.

```
#reading data
data1 <- read_csv("car_sales.csv")
```

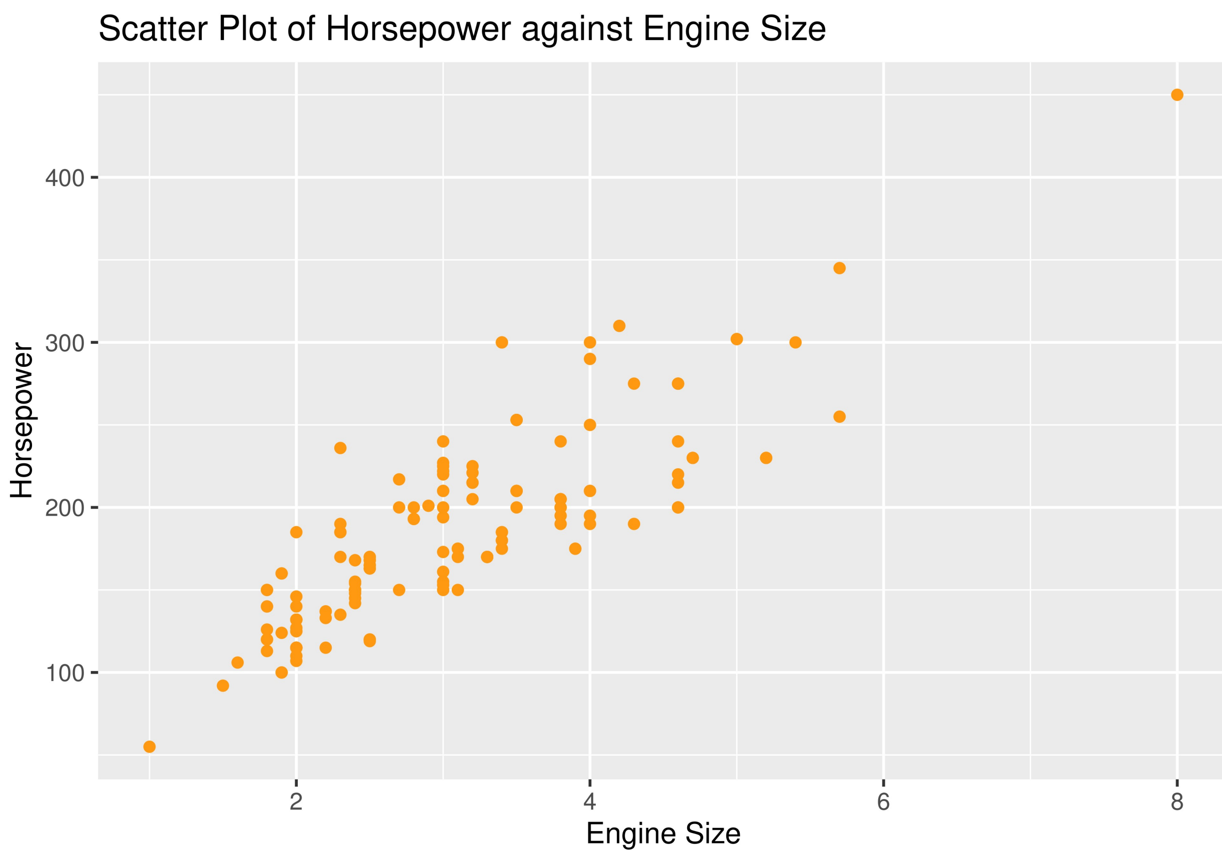
```
## Rows: 152 Columns: 11
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Manufacturer, Model
## dbl (9): price, Engine_size, Horsepower, Wheelbase, Width, Length, Curb_weig...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
data2 <- data1[c('Horsepower','Engine_size')]
data2
```

```
## # A tibble: 152 x 2
##    Horsepower Engine_size
##         <dbl>       <dbl>
##  1         55         1
##  2         92         1.5
##  3        113         1.8
##  4        100         1.9
##  5        106         1.6
##  6        100         1.9
##  7        120         1.8
##  8        120         1.8
##  9        140         1.8
## 10        124         1.9
## # ... with 142 more rows
```

a. Plot horsepower (y axis) against engine size (x axis). Make sure to label your axes. Comment on the form, strength, and direction of the plot. Note if there are any potential outliers.

```r
#Scatter Plot of Horsepower against Engine Size
ggplot(data=data2, mapping = aes(x = Engine_size, y = Horsepower)) +
  geom_point(color='#ff9911') +
  ggtitle("Scatter Plot of Horsepower against Engine Size") +
  xlab("Engine Size") + ylab("Horsepower")
```



Strongly correlated, positive and linear association between horsepower and engine size. Yes, there are few outliers for ex - (8,450)

```
b. Calculate the correlation between horsepower and engine size.
Comment on this value in relation to your scatterplot
```

```r
#Pearson Correlation
pearson <- cor.test(data2$Engine_size, data2$Horsepower, method = "pearson")
pearson
```

```
##
##  Pearson's product-moment correlation
##
## data:  data2$Engine_size and data2$Horsepower
## t = 18.707, df = 150, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7815469 0.8787984
## sample estimates:
```

```
##       cor
## 0.8366494
```

Pearson Correlation = 0.8366494. Value suggests that there is strong correlation which was also seen graphically.

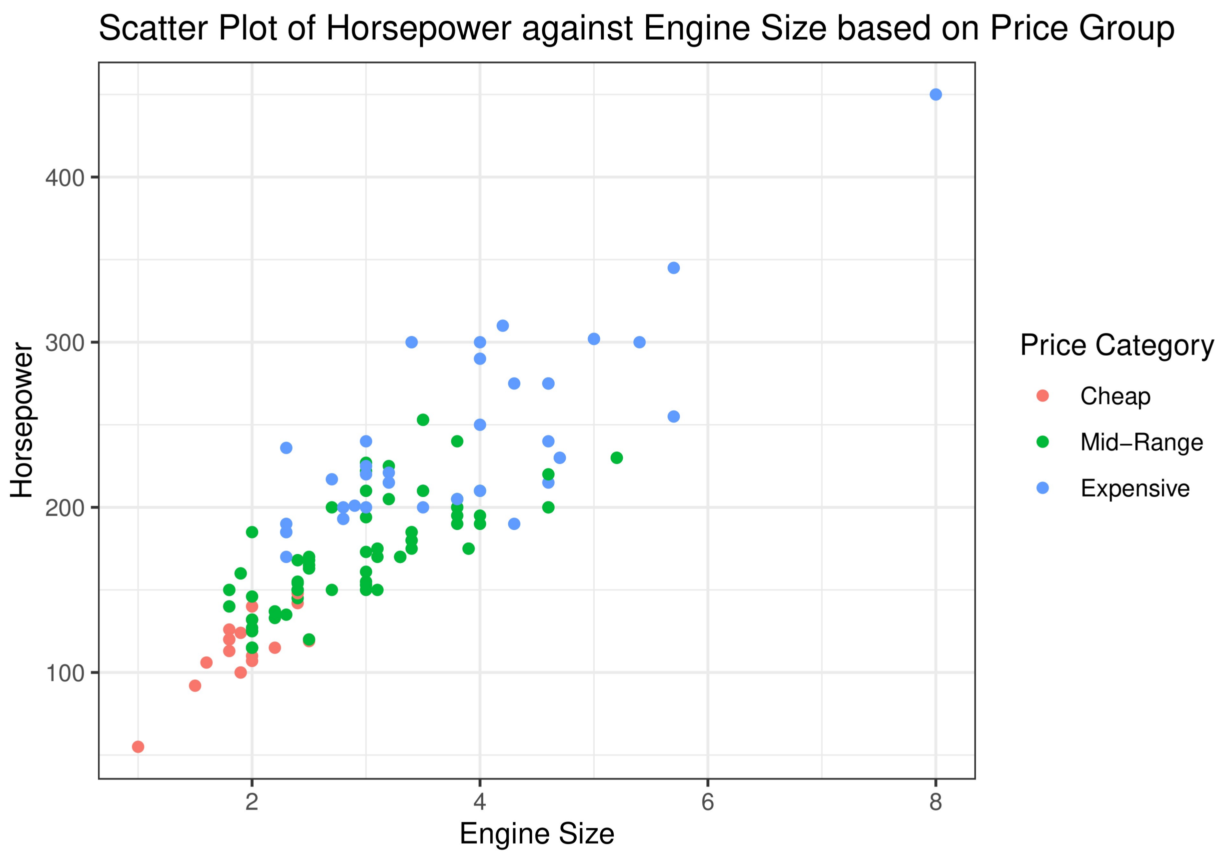
c. Let's break down prices into three groups: the cheapest cars being between 0 and $15000, and mid-range cars being between $15000 and $30000, and the expensive cars costing over $30000. You can use sample code such as this to break price into these three categories.

```
#Breaking price into 3 groups
data1$price_category<- cut(data1$price,breaks=c(0,15000,30000,200000),labels=c("Cheap","Mid-Range","Expe
data1$price_category
```

```
##   [1] Cheap     Cheap     Cheap     Cheap     Cheap     Cheap     Cheap    
##   [8] Cheap     Mid-Range Cheap     Cheap     Cheap     Cheap     Cheap    
##  [15] Cheap     Cheap     Cheap     Mid-Range Mid-Range Cheap     Mid-Range
##  [22] Cheap     Mid-Range Mid-Range Mid-Range Expensive Expensive Mid-Range
##  [29] Mid-Range Mid-Range Mid-Range Mid-Range Mid-Range Mid-Range Mid-Range
##  [36] Mid-Range Mid-Range Mid-Range Mid-Range Expensive Mid-Range Mid-Range
##  [43] Mid-Range Mid-Range Mid-Range Expensive Mid-Range Mid-Range Cheap    
##  [50] Mid-Range Expensive Mid-Range Cheap     Mid-Range Expensive Mid-Range
##  [57] Cheap     Mid-Range Mid-Range Mid-Range Mid-Range Mid-Range Mid-Range
##  [64] Expensive Mid-Range Mid-Range Expensive Expensive Mid-Range Mid-Range
##  [71] Expensive Mid-Range Expensive Mid-Range Mid-Range Mid-Range Mid-Range
##  [78] Mid-Range Mid-Range Mid-Range Expensive Expensive Mid-Range Mid-Range
##  [85] Mid-Range Mid-Range Mid-Range Mid-Range Mid-Range Mid-Range Expensive
##  [92] Mid-Range Mid-Range Mid-Range Mid-Range Mid-Range Mid-Range Mid-Range
##  [99] Mid-Range Expensive Mid-Range Mid-Range Mid-Range Mid-Range Expensive
## [106] Expensive Expensive Expensive Expensive Mid-Range Mid-Range Expensive
## [113] Expensive Mid-Range Expensive Expensive Expensive Mid-Range Expensive
## [120] Mid-Range Expensive Mid-Range Expensive Expensive Expensive Mid-Range
## [127] Mid-Range Mid-Range Mid-Range Mid-Range Mid-Range Expensive Expensive
## [134] Mid-Range Expensive Expensive Expensive Expensive Expensive Mid-Range
## [141] Mid-Range Mid-Range Mid-Range Expensive Mid-Range Mid-Range Expensive
## [148] Expensive Expensive Expensive Expensive Expensive
## Levels: Cheap Mid-Range Expensive
```

d. Plot total horsepower (y axis) against engine size (x axis), but now color points based on which price group they fall into. You can do this by specifying the `col=new_var` option in the `plot()` function. Comment on the results

```
# Scatter Plot of Horsepower against Engine Size based on Price group
ggplot(data=data1, mapping = aes(x = Engine_size, y = Horsepower)) +
  geom_point(aes(color = price_category)) +  theme_bw() +
  ggtitle("Scatter Plot of Horsepower against Engine Size based on Price Group") +
  xlab("Engine Size") + ylab("Horsepower") +labs(colour="Price Category")
```



Comment: It is evident that cheap cars have smaller engine size and horsepower, followed by mid range cars, while expensive cars have most superior engine size and horsepower. It can be observed that around (3,200) mid-range and expensive cars both are present which essentially means that for that engine size and horsepower, mid range cars are available instead of expensive cars.

e. Create a new categorical variable that indicates whether the fuel efficiency is greater than 30. Use the following example code as a template:

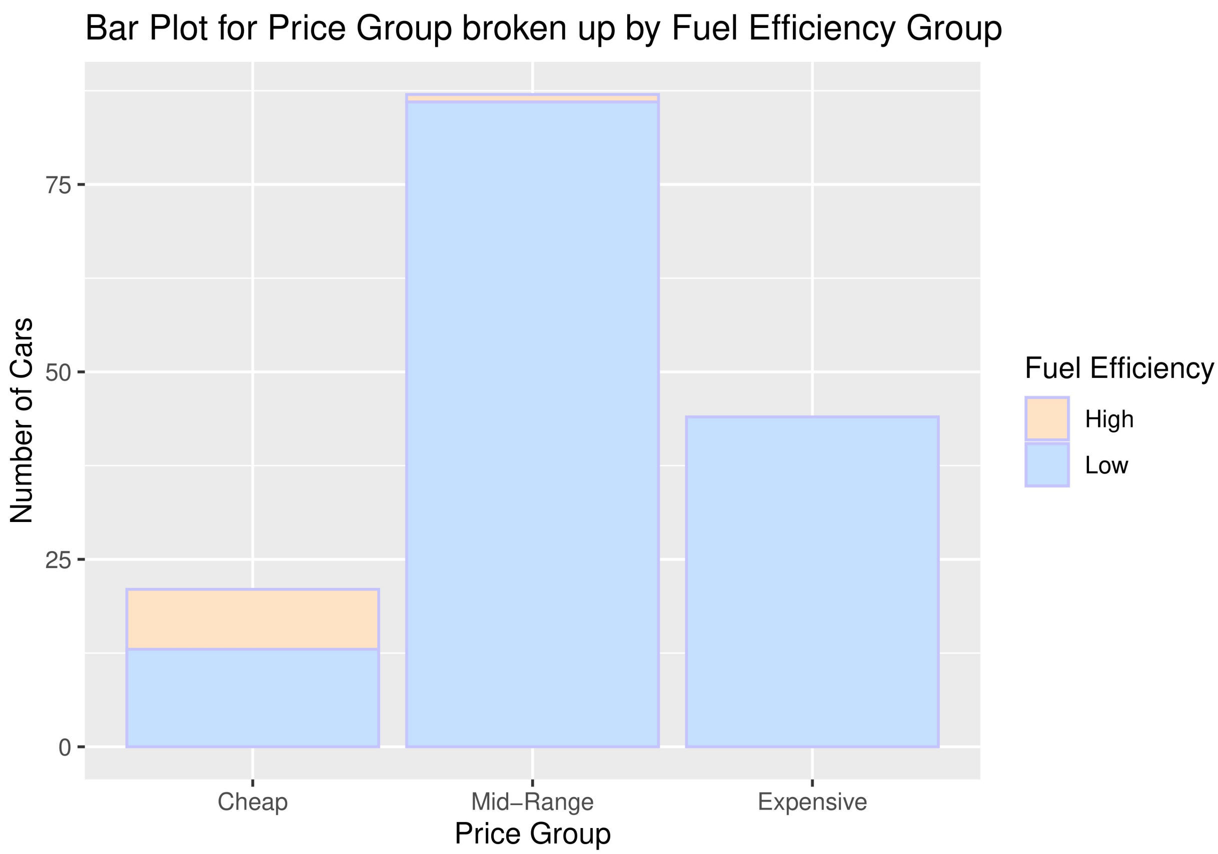
```
#Created new variable fuel_eff_category
data1$fuel_eff_category <- ifelse(data1$Fuel_efficiency > 30, "high_efficient",
                                  "low_efficient")
data1$fuel_eff_category
```

```
##   [1] "high_efficient" "high_efficient" "low_efficient"  "high_efficient"
##   [5] "high_efficient" "high_efficient" "high_efficient" "high_efficient"
##   [9] "high_efficient" "high_efficient" "low_efficient"  "low_efficient"
##  [13] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [17] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [21] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [25] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [29] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [33] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [37] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [41] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [45] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [49] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
```

```
##  [53] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [57] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [61] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [65] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [69] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [73] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [77] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [81] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [85] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [89] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [93] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
##  [97] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
## [101] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
## [105] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
## [109] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
## [113] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
## [117] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
## [121] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
## [125] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
## [129] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
## [133] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
## [137] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
## [141] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
## [145] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
## [149] "low_efficient"  "low_efficient"  "low_efficient"  "low_efficient"
```

f. Create a stacked barplot with a bar for each price group (i.e. use `new_var` from above). Each bar should be broken up into two pieces: one for high fuel efficiency and one for low fuel efficiency. Make sure to label your axes and add a legend. Comment on the results.

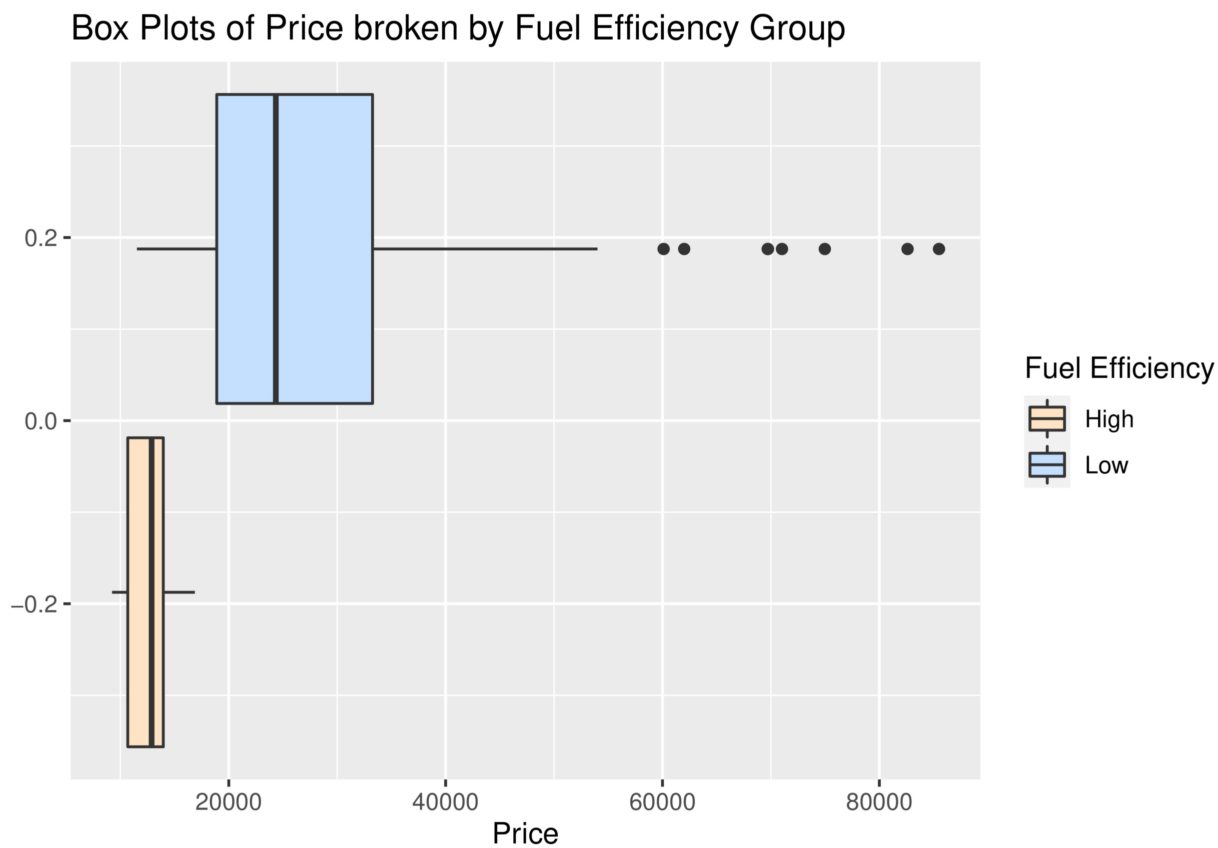
```r
# Bar Plot for Price Group broken up by Fuel Efficiency Group
library(ggplot2)

ggplot(data1, aes(x = price_category, fill = fuel_eff_category)) +  geom_bar() +
  scale_fill_manual(values=c("#ffe4c4", "#c4dfff"), labels = c("High", "Low")) +
  geom_bar(color = "#c7c4ff") +
  guides(fill = guide_legend(title = "Fuel Efficiency")) +
  ggtitle("Bar Plot for Price Group broken up by Fuel Efficiency Group ") +
  xlab("Price Group") + ylab("Number of Cars")
```



Comment: Using the plot we can observe that for cheap cars, there are almost as many as high fuel efficient cars as low fuel efficient cars. In the mid-range section, there are very few high fuel efficient cars while, in expensive price section there are no high fuel efficient cars. It can be inferred that as the price of car increases, fuel efficiency decreases. In real life this trend is followed as expensive cars have bigger engines which require more fuel and hence less efficiency.

g. Make side-by-side boxplots of `price` (not price groups), broken down by fuel efficiency group (low vs. high). Comment on the result:

```
# Box Plots of Price broken by Fuel Efficiency Group
ggplot(data1, aes(x=data1$price, fill= fuel_eff_category)) +
  geom_boxplot() + scale_fill_manual(values=c("#ffe4c4", "#c4dfff"),
                                    labels = c("High", "Low"))  +
  ggtitle("Box Plots of Price broken by Fuel Efficiency Group") + xlab("Price")+
  guides(fill = guide_legend(title = "Fuel Efficiency"))
```



Comment: In this plot it can be observed that there are very few cars in the high efficiency category with respect to low efficiency. Median price of the highly efficient cars lies around 12000 while the median price of low efficient cars lies around 25000. Outliers can be observed.Box-plot of High fuel efficient cars in comparitively smaller than low fuel efficient cars.

Question 2) Probability: PPV and NPV. A test is created to help detect a disease The test is administered to a group of 84 subjects known to have the disease. Of this group, 59 test positive. The test is also administered to a group of 428 subjects known to not have the disease. Of this group, 12 test positive.

```
a. Present this data in a tabular form similar to the following:
```

Test	Have disease	Do not have disease	Total
Positive	59	12	71
Negative	25	416	441
Total	84	428	512

```
#Creating matrix and converting into table
tab <- matrix(c(59, 12, 71, 25, 416, 441, 84, 428, 512), ncol=3, byrow=TRUE)
rownames(tab) <- c('Positive','Negative','Total')
colnames(tab) <- c('Disease','Not Disease','Total')
tab <- as.table(tab)
library(knitr)
    kable(x=tab, digits=2,row.names=T, format="markdown")
```

	Disease	Not Disease	Total
Positive	59	12	71
Negative	25	416	441
Total	84	428	512

b. Calculate the sensitivity and specificity of this test directly from the data.

```
#Calculating Sensitivity
sensitivity = tab[1,1]/(tab[1,1]+tab[2,1])
sensitivity
```

```
## [1] 0.702381
```

0.702381 is the sensitivity

```
#Calculating Specificity
specificity = tab[2,2]/(tab[2,2]+tab[1,2])
specificity
```

```
## [1] 0.9719626
```

0.9719626 is the Specificity

c. Assume that the prevalence of the disease is 2.7%. Calculate the NPV and PPV with this prevalence.

```
#Finding PPV
prevalence = 0.027
PPV = (sensitivity * prevalence)/((sensitivity * prevalence)+
                                    ((1 - specificity) * (1 -prevalence)))
PPV
```

```
## [1] 0.410086
```

PPV is 0.410086

```
#Finding NPV
NPV= (specificity*(1-prevalence))/((((1-sensitivity)*prevalence)+
                                     ((specificity)*(1-prevalence)))
NPV
```

```
## [1] 0.9915747
```

NPV is 0.9915747

d. What conclusions can be drawn regarding the effectiveness of this test?

The chance that a person with a positive test result actually has the disease is 41.0086% which is pretty bad considering almost 60% people that were given positive result, don't actually have disease. And the chance that a person with a negative test result actually does not have the disease is 99.15747%, this is very accurate as only less than 1% people have disease if the report comes negative. The reason for such bad PPV is prevalence value. As prevalence is so small, PPV will be bad. In this case we can't trust on single examination result.

3. Probability: Widget production. Consider a factory that produces widgets. These widgets can have one (or more) of three different types: A, B, and C. Suppose that 20% of these widgets have type A, 40% have type B, 10% have both type A and B, and 50% have type C. Any widget of type C only has one type (i.e., there are no widgets of types A and C, B and C, or A, B, and C). Widgets can either be defective (D) or functional (D^c). Denote by $\Pr(D|X)$ the probability that a widget that has type X is defective. The factory knows that $\Pr(D|A) = 0.25$, $\Pr(D|B) = 0.6$, $\Pr(D|A \cap B) = 0.4$, and $\Pr(D|C) = 0.2$.

a. What is the probability that a widget is defective, $\Pr(D)$? (Hint: Recall the Law of Total Probability.)

3) Law of Total Probability is for mutually exclusive events. That is not the case here.

a) $$Pr(D) = Pr(A)\,Pr\left(\frac{D}{A}\right) + Pr(B)\,Pr\left(\frac{D}{B}\right) + Pr(C)\,Pr(D/C) - Pr(A \cap B) \cdot Pr\left(\frac{D}{A \cap B}\right).$$

$Pr(A) = 0.2$ $\left(20\% = \frac{20}{100} = 0.20 \right)$

$Pr(B) = 0.4$

$Pr(C) = 0.5$

$Pr(A \cap B) = 0.1$.

Given $Pr\left(\frac{D}{A}\right) = 0.25$ $\quad Pr\left(\frac{D}{B}\right) = 0.6$

$Pr\left(\frac{D}{C}\right) = 0.2$ $\quad Pr\left(\frac{D}{A \cap B}\right) = 0.4$

$$Pr(D) = 0.25 \times 0.2 + 0.6 \times 0.4 + 0.5 \times 0.2 - 0.1 \times 0.4$$

$$= 0.05 + 0.24 + 0.1 - 0.04$$

$$= \underline{0.35}.$$

Prob widget is defective; $Pr(D) = 0.35$

b. What is the probability that a defective widget is of type B, or $\Pr(B|D)$?

If $Pr(D) = 0.35$
$Pr(D') = 1 - Pr(D) = 0.65.$

b). $Pr\left(\frac{B}{D}\right) = Pr\left(\frac{D}{B}\right) \cdot \frac{Pr(B)}{Pr(D)}.$

(Bayes Theorem)

$= 0.6 \times \frac{0.4}{0.35.} = \underline{0.6857}.$

d. Prob that defective widget is of type B $Pr\left(\frac{B}{D}\right) = 0.6857$

c. What is the probability that a non-defective (i.e., functional) widget is either type A or type B (or both), i.e., what is $\Pr(A \cup B|D^c)$?

c). To find $Pr\left(\frac{A \cup B}{D^c}\right)$

$= \frac{P_r(A \cup B)}{P_r(D^c)} \cdot Pr\left(\frac{D^c}{A \cup B}\right)$

(Bayes Theorem).

$P_r(A \cup B) = P(A) + P(B) - P(A \cap B)$
$= 0.2 + 0.4 - 0.1 = 0.5$

Now,

$$P\left(\frac{D'}{A \cup B}\right) = 1 - P\left(\frac{D}{A \cup B}\right)$$

$$= 1 - \left(\frac{P(D \cap A) + P(D \cap B) - P(D \cap A \cap B)}{P(A \cup B)}\right)$$

Now $P(D \cap A) = P\left(\frac{D}{A}\right) P(A)$

$= 0.25 \times 0.2 = 0.05$

$P(D \cap B) = P\left(\frac{D}{B}\right) P(B) = 0.4 \times 0.6$

$= 0.24$

$P(D \cap A \cap B) = P\left(\frac{D}{A \cap B}\right) \cdot P(A \cap B)$

$= 0.4 \times 0.1 = 0.04$

$$P\left(\frac{D'}{A \cup B}\right) = 1 - \left[\frac{0.05 + 0.24 - 0.04}{0.5}\right]$$

$$= 1 - \frac{0.25}{0.5} = \underline{0.5}$$

$$P\left(\frac{A \cup B}{D^c}\right) = \frac{P(A \cup B) \cdot P\left(\frac{D'}{A \cup B}\right)}{P(D')}$$

$$= \frac{0.5 \times 0.5}{0.65} = 0.384615$$

Probability that a non defective widget is either A or B or Both is

$$Pr\left(\frac{A \cup B}{D^c}\right) = 0.384615$$

4. Probability: Inclusion-exclusion. Recall that the additive rule tells us for events A and B that are not mutually exclusive that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. We can extend this additive rule to more than two events, which gives us the general inclusion-exclusion identity as follows:

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + ... + (-1)^{n+1} P(A_1 \cap A_2 \cap ... \cap A_n)$$

a. Explicitly write the inclusion-exclusion identity for $n = 3$ events, A_1, A_2, A_3 (i.e., reduce down so that there aren't summations).

Answer-4

a) $P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \cdots$

$\cdots + (-1)^{n+1} P(A_1 \cap A_2 \cdots \cap A_n)$.

For $n=3$.

$P(\cup_{i=1}^{3} A_i) = \sum_{i=1}^{3} P(A_i) - \sum_{1 \le i < j \le 3} P(A_i \cap A_j) + \sum_{1 \le i < j < k \le 3} P(A_i \cap A_j \cap A_k)$

$P(A \cup B \cup C) = \underbrace{\sum_{i=1}^{3} P(A_i)}_{\text{Part 1}} - \underbrace{\sum_{1 \le i < j \le 3} P(A_i \cap A_j)}_{\text{Part 2}} + \underbrace{\sum_{1 \le i < j < k \le 3} P(A_i \cap A_j \cap A_k)}_{\text{Part 3}}$

Part-1.

$\sum_{i=1}^{3} P(A_i) = P(A) + P(B) + P(C)$.

Part-2.

$\sum_{1 \le i \le j \le 3} P(A_i \cap A_j) = P(A \cap B) + P(B \cap C) + P(A \cap C)$

Part-3.

$\sum_{1 \le i < j < k \le 3} P(A_i \cap A_j \cap A_k) = P(A \cap B \cap C)$.

$\therefore P(A \cup B \cup C) = P(A) + P(B) + P(C)$

$- P(A \cap B) - P(B \cap C) - P(A \cap C)$.

$+ P(A \cap B \cap C)$.

Alternatively,

for $n=2$ we know

$$P(A\cup B) = P(A) + P(B) - P(A\cap B)$$

we have to find $P(A\cup B\cup C)$

Now assume $P(A\cup B) = P(D)$.

$\therefore$ $P(D) = P(A) + P(B) - P(A\cap B)$

so, $P(D\cup C) = P(D) + P(C) - P(D\cap C)$

Substituting $P(A\cup B)$ in place of $P(D)$

$$P(A\cup B) = P(A) + P(B) - P(A\cap B) + P(C) - P(A\cap C) - P(B\cap C) + P(A\cap B\cap C).$$

b. Suppose an integer from 1 to 1000 (inclusive) is chosen uniformly at random (i.e., with equal probability). What is the probability that the integer is divisible by 5, 7, or 13?

b) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$

To find : prob of integer is divisible by 5, 7 or 13.

$1 \leq n \leq 1000$, Let $A = 5$, $B = 7$ & $C = 13$

Step-1:

Using AP we find number of int divisible by 5 in first 1000 natural numbers

$1000 = 5 + (n-1)5$

$n = 199 + 1 = \underline{200}$

$P(A) = \frac{200}{1000} = \underline{0.2}$

Similarly for 7 & 13

$P(B) = \frac{142}{1000} = 0.142$

$P(C) = \frac{76}{1000} = 0.076$

Step - 2:

Now for $P(A \cap B)$

we take LCM of A & B

$LCM(5, 7) = 35 = \underline{A \cap B}$

so $P(A \cap B) = P(35) =$

Similarly using AP.

$1000 = 35 + (n-1)35$

$n = 28.57 = \underline{28}$

$P(A \cap B) = P(35) = \frac{28}{1000} = 0.028$

Similarly for $P(B \cap C)$ & $P(C \cap A)$

$P(B \cap C) =$ LCM of 7 & 13 $\Rightarrow 91$

$P(B \cap C) = \frac{10}{1000} = 0.01$

$P(C \cap A) =$ LCM of 5 & 13 $\Rightarrow 65$

$P(C \cap A) = \frac{15}{1000} = 0.015.$

Step-3

$P(A \cap B \cap C) \Rightarrow$ LCM of 5, 7 & 13

$= 455$

$P(A \cap B \cap C) = \frac{2}{1000} = 0.002.$

substitute values into formula

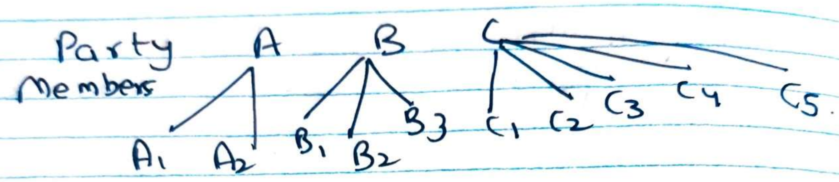
$= 0.2 + 0.142 + 0.076 - 0.028$
$- 0.010 - 0.015 + 0.002$
$= \underline{0.367}$

Prob that the integer is divisible by 5, 7 or 13 is $\underline{0.367}$.

5. Combinatorics: Consider a political setting where there are three political parties, A, B, and C vying for seats on a 3-person committee. Party A has 2 members, B has 3 members, and C has 5 members. Members of parties are distinguishable from each other, but positions on the committee are indistinguishable from each other.

a. How many ways are there of forming an unordered 3-person committee?

As).



a) Unordered 3-person comittee.

$n = 10 \quad r = 3$

$$\text{ways} = \frac{n!}{r!(n-r)!} = \frac{10!}{7! \cdot 3!} = \underline{120}$$

120 ways are there.

b. How many different party breakdowns (e.g., ABC, CCC, etc.) are possible when forming an unordered 3-person comittee?

b). Case -1

All diff : $\underline{A}$ $\underline{B}$ $\underline{C}$ => 1 way.

Case-2 2 member of same party

$\underline{A}$ $\underline{A}$ _ => 2 ways

$\underline{B}$ $\underline{B}$ _ => 2 ways

$\underline{C}$ $\underline{C}$ _ => 2 ways

Case-3 3 members of same party

$\underline{B}$ $\underline{B}$ $\underline{B}$ & $\underline{C}$ $\underline{C}$ $\underline{C}$ => 2 ways

Not applicable for Party - A

Total ways = 9 //

c. How many ways are there of forming an unordered 3-person committee if at least one member must be from party A?

c) Total cases = $^{10}C_3$

No member of Party A cases

= $^{8}C_3$

(members in B + members in C = 8)

For Atleast one member of party A.

= $^{10}C_3 - {}^{8}C_3$

= 120 - 56 = $\underline{64}$

There are 64 ways.

6. Combinatorics: Miscellaneous counting.

   a. There are 20 indistinguishable children who would like to have one ice cream cone each. There are 6 distinct flavors of ice cream. How many distinct collections of ice cream cones are there where at least two children must order each flavor?

A6)

a) Let each dash _ represent children & F_1 = icecream flavor-1, similarly F_2, F_3, F_4, F_5 & F_6.

20 ~~was~~ indistinguishable children

_ _ _ _ _ _ _ _ _ _

_ _ _ _ _ _ _ _ _ _

Atleast 2 children must order each flavour, so lets say C_1 & C_2 order F_1.

Similarly C_3 & C_4 order F_2 ...

so on till F_6.

$\boxed{F_1 \; F_1}$ $\boxed{F_2 \; F_2}$ $\boxed{F_3 \; F_3}$ $\boxed{F_4 \; F_4}$ $\boxed{F_5 \; F_5}$

$\boxed{F_6 \; F_6}$ _ _ _ _ _ _ _ _

Now there are 8 children who can chose from 6 flavours

Applying Stars & Bars concept.

For non negative constraints

$$^{n+k-1}C_{k-1}$$

where

$k = 6$ flavours of ice cream

$n = 8$ number of children

$$^{6+8-1}C_{6-1} = {}^{13}C_5$$

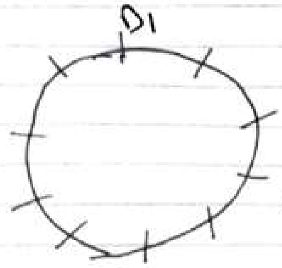
$$= \frac{13!}{8!\,5!} = 1287$$

There are 1287 ways

b. There are five cats and five dogs, all distinguishable from one another. How many distinct ways are there of seating them at a round table such that every cat is adjacent to two dogs and every dog is adjacent to two cats? Note that here two orderings are not considered distinct if it is possible to rotate one and achieve the other. For instance, if there are only four seats at the table, the order Cat 1 - Dog 1 - Cat 2 - Dog 2 is the same as Cat 2 - Dog 2 - Cat 1 - Dog 1.

b) Let there be a circular table
D_1, D_2, D_3, D_4, D_5 be dogs

C_1, C_2, C_3, C_4, C_5 be cats



In order to arrange cats and dogs, we place D_1 at top, which will be starting position.

Now there are 4 dogs and 4 positions.
ways = $4! = \underline{24}$

For cats there are 5 positions and 5 cats
ways = $5! = \underline{120}$

Total ways = $4!\,5!$

$= 24 \times 120$
$= \underline{2880}$.

There are 2880 ways

Short Answers:

- About how long did this assignment take you? Did you feel it was too long, too short, or reasonable?

9-10 hours including reading slidesz.

- Who, if anyone, did you work with on this assignment?

No one

- What questions do you have relating to any of the material we have covered so far in class?

Visual interpretation, correlation, sensitivity, specificity, PPV, NPV, Probability and Combinatorics.