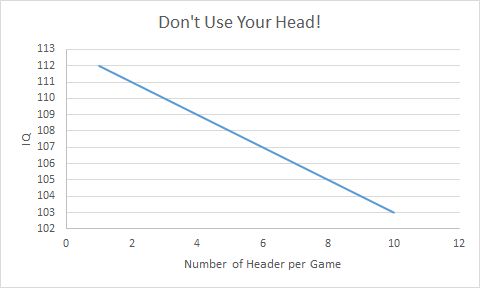
Brent Cottrell, Curt Winegar, Deepti Chevvuri, Jacob McCracken, Jessi Ritcher

STAT 526 Homework 1

1. We are going to use applications that Google offers as our main form of communication. In our team google site we have a to-do list that includes dates of when items need to be completed and who is doing what. We plan on meeting at 4pm Central Time every Sunday. Each group member will complete the homework assignments on their own first and then we will meet on our scheduled time to go over what each of have come up with. We feel this the best way to learn as individuals as well as a group. Jacob McCracken will be the Project Lead. The Administrative Lead will direct the flow of the meetings from everyone getting a chance to present their information and then going through each section of an assignment and coming up with an agreeable answer. The Administrative Lead will also make the meeting schedule. Since we have five in our group a majority rule will be in place when we come to a crossroads. As part of an application with Google will use the calendar to schedule weekly meetings and others on a need-to-meet basis.

2.

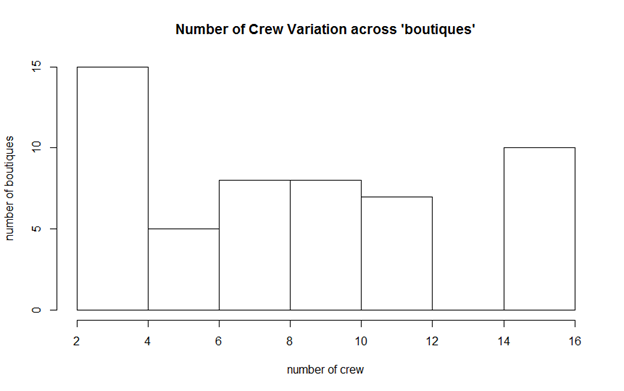
1. The study population of interest in this case are all soccer players.
2. The sample from this population is 60 males ages 14-29 that play soccer at least five times a week.
3. Two variable of interest- IQ (Response Variable), group of players grouped according to number of average headers a game (Explanatory Variable).
4. The variable IQ is Quantitative, whereas the groups are Qualitative or categorical. Because it is not number of people with IQs 103 or 112, they are group of players.
5. The statistics describes the variation in the Average IQ of players along with the number of times they head the ball on average per game.
6. The parameters related to this study would be all soccer players, and not just the ones used in the sample. Based on this study we can see that male soccer players that play regularly (five times a week) between the ages of 14-29 and that head the ball often (10 or more times per game) are more likely to have a lower IQ score than their teammates of the same demographic that head the ball one or fewer times per game.
7. A graph plotted between IQ and Number of headers per game, would look like this

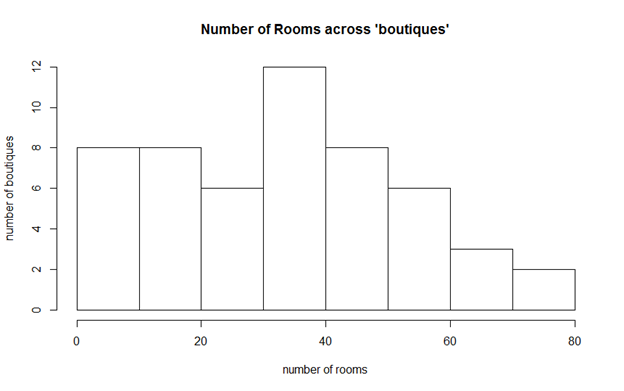


There is a direct relationship with a lower IQ based on the more headers one performs in a game. The response variable in this case(IQ), decreases with an increase in the Explanatory variable (average number of headers per game). The relationship is linear (form), with negative association (negative slope or direction).

3.

1. All the boutique hotels nationwide are the population of interest.
2. The sample in this example are the 53 boutique hotels effectively managed in 2014 that were used in the study.
3. The two variables are: - Number of occupied rooms (Explanatory Variable), number of crew members (Response Variable)
4. The above variables are quantitative as they both deal with number for cleaning crews and occupied rooms.
5. Knowing the mean for each of the variables in the sample would be a good indicator of the ratio needed to keep customers happy at the most efficient level. The statistics of our interest is the variation of the crew members (who are required to clean the occupied rooms) with the variation in the number of occupied rooms. Also knowing the standard deviation and a confidence interval will help us make a more accurate decision based on the data.





**R Code:**

#3.f

> #Import cleaningoriginal.csv

> cleaningOriginal<-read.table("G:/MIS/Sem 2/Stat 526/Homework 1/CleaningOriginal.csv",header=T,sep=",",quote="")

>

> #3.g

> #create histogram of number of crews

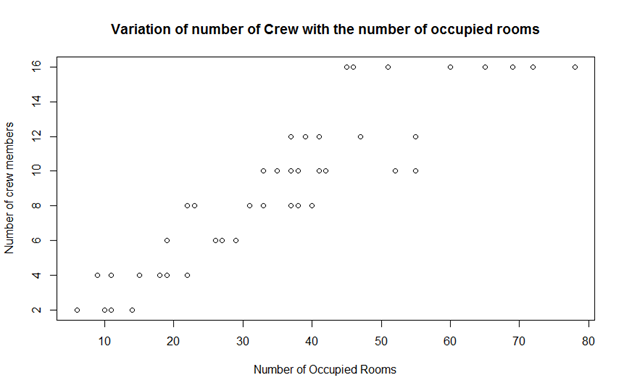
> hist(cleaningOriginal$NumberOfCrews,main="Number of Crew Variation accross 'boutiques'", xlab="Number of Crew", ylab="Number of boutiques")

>

> # create histogram of number of occupied rooms

> hist(cleaningOriginal$NumberOfRooms,main="Number of Occupied Rooms Variation accross 'boutiques'", xlab="Number of Occupied Rooms", ylab="Number of boutiques")

1. The histograms only show the variation of number of crew, number of occupied rooms separately across different boutiques but they do not provide a direct relation between the variation of Response Variable (number of crew members) and Explanatory Variable (Number of rooms occupied).
2. We can see a trend that as the number of rooms increases the number of crews increases as well. Before, we couldn’t see this relationship. A graph is plotted to show the variation of number of crew across the variation of number of occupied rooms. The graph shows a moderately strong, positive linear association i.e. the number of crew members is found to increase with the increase in the number of occupied reasons, this graph will help us to approximately predict the number of crew members that are required to assist a given number of occupied rooms.

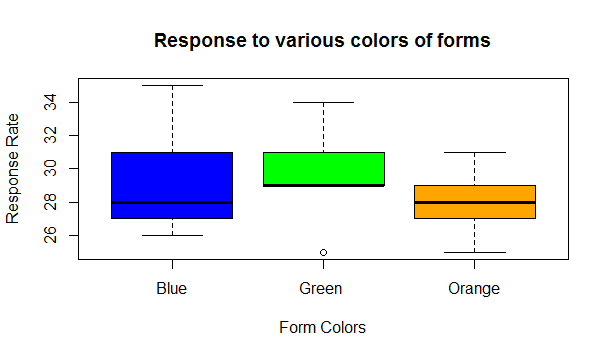


**R Code:**

> #3.h

> # Relational Plot between number of occupied rooms and number of crew

> plot(CleaningOriginal$NumberOfRooms,CleaningOriginal$NumberOfCrews,main="Variation of number of Crew withe the number of occupied rooms", xlab="Number of Occupied Rooms", ylab="Number of crew members")

4.

1. **R Code:**

#Import cleaningoriginal.csv

> responseToColor<-read.table("G:/MIS/Sem 2/Stat 526/Homework 1/response.csv",header=T,sep=",",quote="")

>

> #4.a,b

> #cretae a histogram to display the relation between the color and number of responses

> plot(responseToColor$Color,responseToColor$Response.Rate,col = c('blue', 'green','orange'),ylab="Response Rate", xlab="Form Colors", main="Response to various colors of forms")

1. Description of the Graphical Display:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Color** | **Maximum** | **Upper Quartile** | **Median** | **Lower Quartile** | **Minimum** | **Outlier**  **(if any)** |
| **Blue** | 35 | 31 | 28 | 27 | 26 | - |
| **Green** | 34 | 31 | 29 | 29 | 29 | One less than 3/2 times of the lower quartile. |
| **Orange** | 31 | 29 | 29 | 27 | 25 | - |

Maximum- is the greatest value excluding the outlier.

Minimum- is the least value excluding the outlier.

Upper Quartile- 25% of the data is greater than this value

Median- 50% of the data is less/ greater than this value

Lower Quartile- 25% of the data is less than this value

Analysis-

Mean of Green is higher compared to the mean of the other colors, at the same time variation in the response rate (across various places) of green is lesser that that of blue. Though the variation in response rate of orange is similar to green, the values of Response rates of green are higher. From this analysis we can conclude that on a whole, Green forms were better accepted compared to blue or orange forms.

Note: the effect of the outlier for green has not been considered for this analysis.