Homework2

Steven Black

2/12/2019

## Homework 2 for IS677 Intro to Data Science

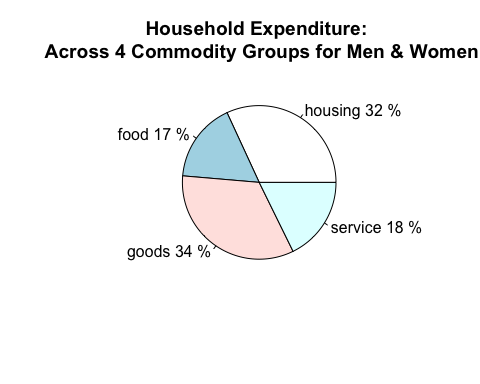
1. The household data is part of a data set collected from a survey of household expenditure and give the expenditure of 20 single men and 20 single women on four commodity groups. The units of expenditure are Hong Kong dollars, and the four commodity groups are as follows:

* housing: housing, including fuel and light,
* food: foodstuff, including alcohol and tobacco,
* goods: other goods, including clothing, footwear and durable goods,
* service: services, including transport and vehicles.

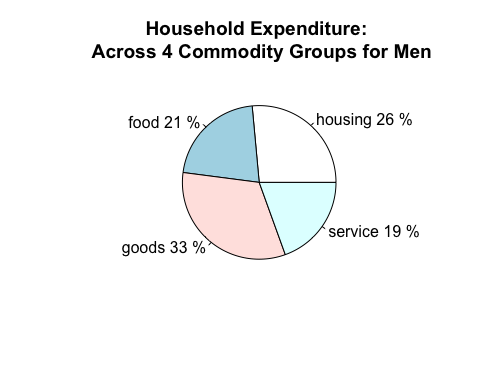
The aim of the survey was to investigate how the division of household expenditure between the four commodity groups depends on total expenditure and to find out whether this relationship differs for men and women. Load the data into a data frame in R and use appropriate graphical methods to answer these questions and state your conclusions.

household <- read.table("household.txt", header = TRUE, sep = ",")

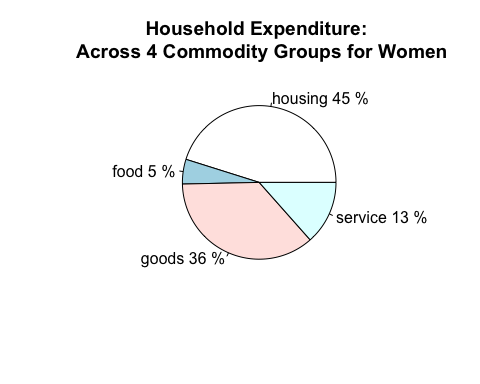
group\_sums <- colSums(household[,1:4])  
pct <- round(group\_sums/sum(group\_sums)\*100)  
lbls <- paste(labels(group\_sums), pct, "%")  
pie(group\_sums,  
 main = "Household Expenditure: \n Across 4 Commodity Groups for Men & Women",  
 labels = lbls)



groups\_sums\_grouped <- aggregate(. ~ household$gender, household, sum)  
group\_sums\_men <- as.numeric(groups\_sums\_grouped[2, 2:5])  
pct\_men <- round(group\_sums\_men/sum(group\_sums\_men)\*100)  
lbls\_men <- paste(labels(group\_sums), pct\_men, "%")  
pie(group\_sums\_men,  
 main = "Household Expenditure: \n Across 4 Commodity Groups for Men",  
 labels = lbls\_men)



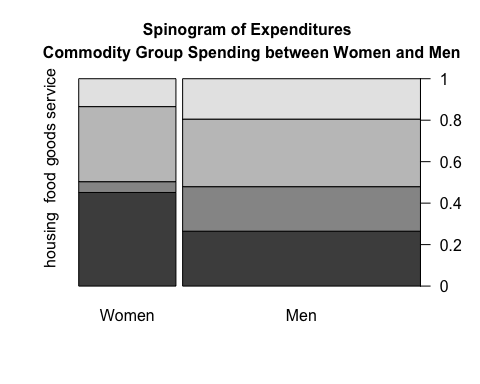
group\_sums\_women <- as.numeric(groups\_sums\_grouped[1, 2:5])  
pct\_women <- round(group\_sums\_women/sum(group\_sums\_women)\*100)  
lbls\_women <- paste(labels(group\_sums), pct\_women, "%")  
pie(group\_sums\_women,  
 main = "Household Expenditure: \n Across 4 Commodity Groups for Women",  
 labels = lbls\_women)



library(vcd)

## Loading required package: grid

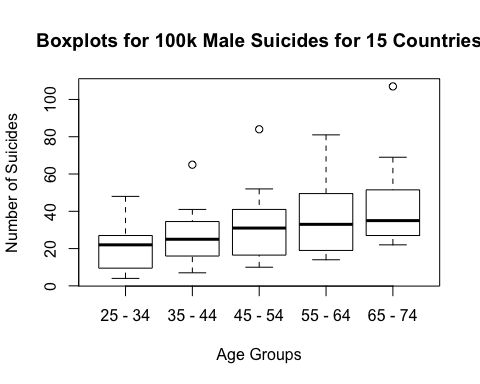
gsg <- groups\_sums\_grouped[2:5]  
row.names(gsg) <- c("Women", "Men")  
spine(as.matrix(gsg), main = "Spinogram of Expenditures \n Commodity Group Spending between Women and Men")

 The pie charts and the spinogram above clearly show there is a difference in spending habits between men and woman. Men spend roughly the same proportion of money on each of the four categories while women spend much more on housing and goods. It’s possible that women like remodeling their kitchens and lots of clothes and so spend more money in the housing and goods categories. Another explanation is that men eat a lot and are lazy so they spend more in the food and service categories. Explanations of this type rarely come down to one factor and so it is likely that the true cause is a little of both and possibly other factors.

1. Mortality rates per 100,000 from male suicides for a number of age groups and a number of countries are provided in the suicides2 data set. Load the data into R and construct side-by-side box plots for the data from different age groups. Comment on what the graphic tells us about the data.

suicides <- read.table("suicides.txt", header = TRUE, sep = ",")

num\_rows <- nrow(suicides)  
boxplot(suicides,  
 main = paste("Boxplots for 100k Male Suicides for", toString(num\_rows), "Countries"),  
 xlab = "Age Groups",  
 ylab = "Number of Suicides",  
 names = c("25 - 34", "35 - 44", "45 - 54", "55 - 64", "65 - 74"))

 Oddly we see more suicides the older men get. There should be more younger men, by the logic that older men must have been younger men at some point, so if the rate of suicides across age groups is constant then there should be a greater number of suicides for younger men. Also strange is the greater variance in suicides between in the groups 25 - 34 and 55 - 64. These age groups roughly correlate to the quarter-life and mid-life chrisis. An explanation for this phenomenon could be that countries with greater pressures for men to succeed see high rates of suicides at these junctures while countries that have less pressure see normal amounts of suicide.