

Quiz - c1 - w1

Fábio Rocha Campos

12 de outubro de 2017

What is the correct mean and standard deviation of the quantity of pasta purchased by time unit by household?

```
q1 = c(mean(df$PASTA), sd(df$PASTA))
q1
```

```
## [1] 1.841545 1.025911
```

In which area are located (i) the poorest household and (ii) the wealthiest household?

```
order.scores = order(df$INCOME)
df = df[order.scores,]
df$rank = rank(df$INCOME)
head(df, n = 1)
```

```
##      HHID TIME    PASTA EXPOS      AGE  INCOME AREA rank
## 35241 1763     1 2.332987      0 32.21623 609.1072    2 10.5
```

```
tail(df, n = 1)
```

```
##      HHID TIME    PASTA EXPOS      AGE INCOME AREA    rank
## 19240  962   20 3.201276      0 28.54533 141067    5 39990.5
```

What is the maximum pasta quantity a household has bought over the whole time period? (Sum the quantity of pasta by household over time and indicate the maximum)

```
q3 = df %>% group_by(HHID) %>% summarize(pasta_qty = sum(PASTA) )
max(q3$pasta_qty)
```

```
## [1] 55.36193
```

What is the average income of households living in area 4?

```
q4 = subset(df, AREA == 4, select = c(HHID, INCOME))
q41 = distinct(q4, HHID, .keep_all = TRUE)
mean(q41$INCOME)
```

```
## [1] 29260.13
```

How many households live in area 2, earn more than 20k, and have purchased more than 30 units of pasta over the whole time period?

```
q5 = df %>% filter(AREA == 2) %>% filter(INCOME > 20000) %>%  
  group_by(HHID) %>%  
  summarize(total_qty = sum(PASTA) ) %>% filter(total_qty > 30)  
nrow(q5)
```

```
## [1] 218
```

What is the correlation between the purchases of pasta and the exposures?

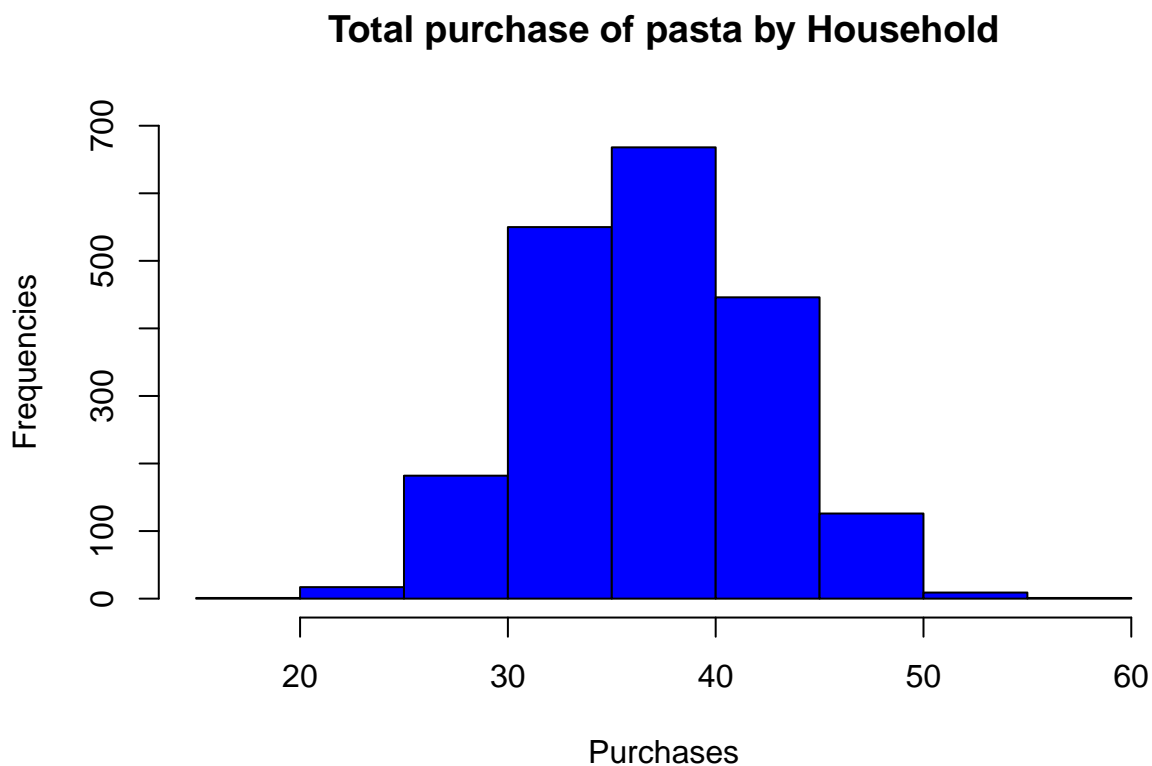
```
cor(df$PASTA, df$EXPOS)
```

```
## [1] 0.3266174
```

Which of the following graphs reports the correct histogram by household of the total purchase of pasta made by the household over the whole period? (Sum the purchases by household and make a histogram.)

Note that the color or exact representation may be different in your version.

```
q7 = df %>% group_by(HHID) %>%  
  summarize(total = sum(PASTA))  
q71 = distinct(q7, HHID, .keep_all = TRUE)  
hist(q71$total, ylim = c(0,700), main = "Total purchase of pasta by Household", xlab = "Purchases", ylab = "Frequencies")
```



Which of the following graphs reports the correct time series of the overall total purchase of pasta? (Sum the purchases by time units and plot the quantity by time unit.)

```
q8 = df %>% group_by(TIME) %>% summarize( total_qty = sum(PASTA))  
plot(q8$total_qty, main = "Time series of overall total purchase of pasta", col = "blue", xlab = "Time"
```

