Total liters of pure alcohol based on continents

Chi-square Test of Association

Phalgun Haribabu Chintal, s3702107 and Syed Junaid Ahmed, s3731300

Last updated: 28 October, 2018

RPubs link information

- Rpubs link: http://rpubs.com/Phalgunhc/434132

Introduction

- This investigation aims to understand between the total liters of pure alcohol serves and continents.
- Every continent serves unique volume of pure alcohol.
- Chi-square Test of Association is chosen for hypothesis testing.
- The main objective is to find out which continent serves more volume of pure alcohol.

Problem Statement

- Is there any association between the total liters of pure alcohol and the continents?
- Total liters of pure alcohol is an important variable to check the volume of pure alcohol served. The continents contain all their respective countries.
- Can this prove which continent serves pure alcohol?
- Do they have a significant association between them?
- Chi-square Test of Association is the best solution to overcome these problems to determine whether there is a significant association between two variables.

Data

- The dataset contains information about the total liters of pure alcohol among continents. Drinks in this dataset are beer, spirit, and wine.
- This dataset was collected from the kaggle website (https://www.kaggle.com/navneethc/drinks)

Data Cont.

- The dataset consists of 6 variables and 193 observations.
- Country: It shows the name of the country that serves pure alcohol.
- beer_servings: the volume of beer served.
- spirit_servings: the volume of spirit served.
- wine_servings: the volume of wine served.
- total_liters_of_pure_alcohol: total liters of pure alcohol that country serves.
- continent: it shows which continent that country belongs.
- The continents in this dataset were AS, AF, EU, NA, SA, and OC. The observation NA (North America) was renamed to NAM because when the tasks are processed NA was treated as Not Available(NA).

Descriptive Statistics and Visualisation

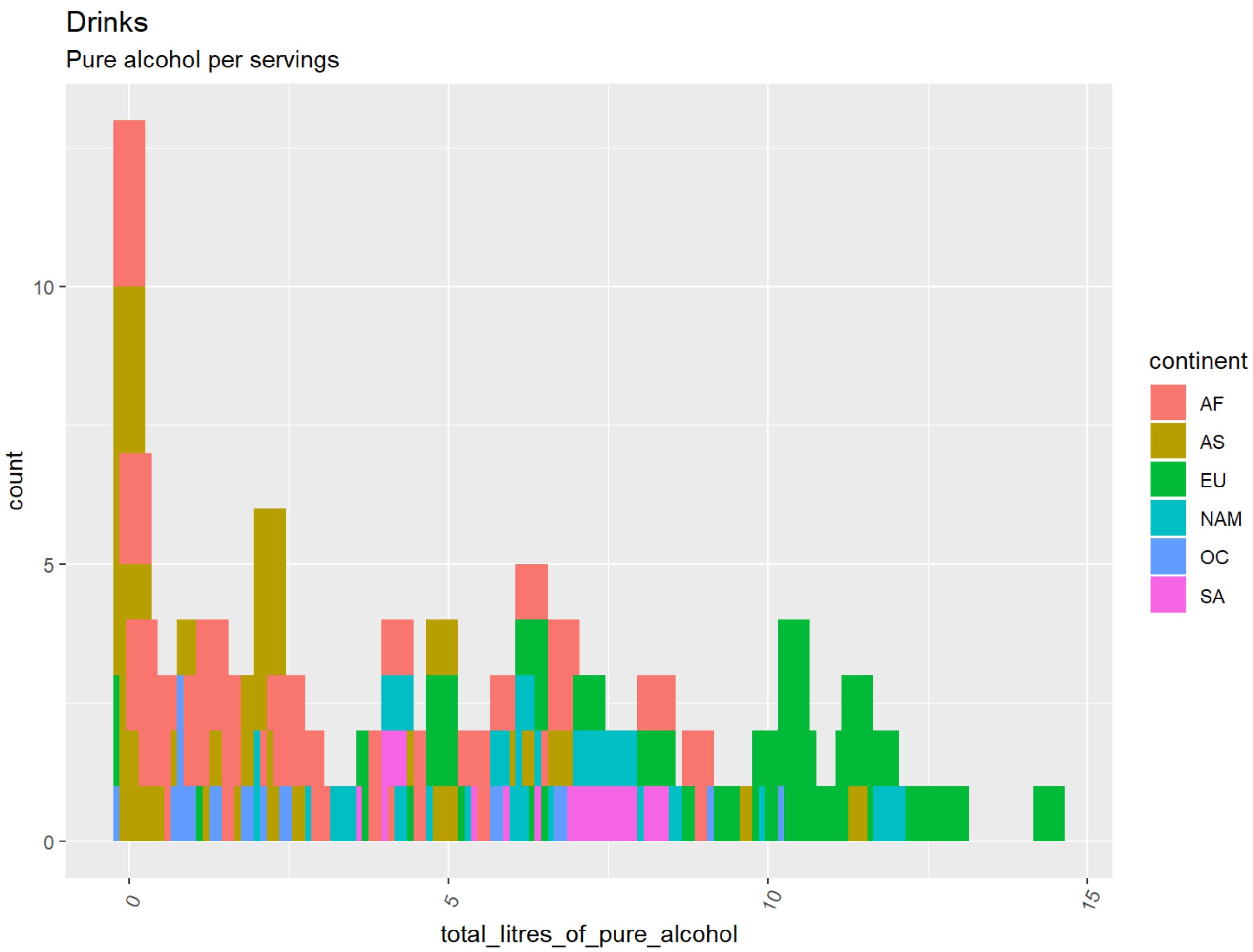
- Descriptive statistics of drinks are grouped by continents.
- This dataset had no missing values nor outliers.

```
drinks <- read.csv("drinks.csv")
table1 <- drinks %>% group_by(continent) %>%
  summarise(Min = min(total_litres_of_pure_alcohol,na.rm = TRUE),
  Q1 = quantile(total_litres_of_pure_alcohol,probs = .25,na.rm = TRUE),
  Median = median(total_litres_of_pure_alcohol, na.rm = TRUE),
  Q3 = quantile(total_litres_of_pure_alcohol,probs = .75,na.rm = TRUE),
  Max = max(total_litres_of_pure_alcohol,na.rm = TRUE),
  Mean = mean(total_litres_of_pure_alcohol, na.rm = TRUE),
  SD = sd(total_litres_of_pure_alcohol, na.rm = TRUE), n = n(),
  Missing = sum(is.na(total_litres_of_pure_alcohol)))
knitr::kable(table1)
```

continent	Min	Q1	Median	Q3	Max	Mean	SD	n	Missing
AF	0.0	0.70	2.30	4.700	9.1	3.007547	2.647557	53	0
AS	0.0	0.10	1.20	2.425	11.5	2.170454	2.770239	44	0
EU	0.0	6.60	10.00	10.900	14.4	8.617778	3.358455	45	0
NAM	2.2	4.30	6.30	7.000	11.9	5.995652	2.409353	23	0
OC	0.0	1.00	1.75	6.150	10.4	3.381250	3.345687	16	0
SA	3.8	5.25	6.85	7.375	8.3	6.308333	1.531166	12	0

Descriptive Statistics Cont.

```r
g <- ggplot(drinks, aes(total_litres_of_pure_alcohol))
g + geom_bar(aes(fill=continent), width = 0.5) +  theme(axis.text.x = element_text(angle=65, vjust=0.6)) +
  labs(title="Drinks", subtitle="Pure alcohol per servings")
```



Hypothesis Testing

- Chi-square Test of Association is used in this investigation.
- H0 : There is no association in the total_litres_of_pure_alcohol between the continents (independence)
- HA: There is an association in the total_litres_of_pure_alcohol between the continents (dependence)
- Decision Rules:
- Reject H0, if p-value < 0.05
- Otherwise, fail to reject H0
- Conclusion:
- If we reject H0, then it is statistically significant.
- Otherwise, it is not statistically significant.

Hypothesis Testing Cont.

```
chi_test <- chisq.test(table(drinks$continent, drinks$total_litres_of_pure_alcohol))
chi_test
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(drinks$continent, drinks$total_litres_of_pure_alcohol)
## X-squared = 546.34, df = 445, p-value = 0.0007097
```

```
pchisq(q=546.34, df = 445, lower.tail = FALSE)
```

```
## [1] 0.000709786
```

Hypothesis Testing Cont.

```
chi_test$observed
```

```
##      
##        0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1 1.2 1.3 1.4 1.5 1.7 1.8
##   AF   3   2   2   1   1   2   1   2   1   0 0   2   0   3   0   1   2   3
##   AS   7   5   2   2   1   1   1   1   0   1 1   0   0   0   1   1   0   0
##   EU   2   0   0   0   0   0   0   0   0   0 0   0   0   1   0   0   0   0
##   NAM  0   0   0   0   0   0   0   0   0   0 0   0   0   0   0   0   0   0
##   OC   1   0   0   0   0   0   0   0   0   1 3   1   1   0   0   1   0   0
##   SA   0   0   0   0   0   0   0   0   0   0 0   0   0   0   0   0   0   0
##      
##        1.9 2 2.2 2.3 2.4 2.5 2.6 2.8 3 3.1 3.4 3.5 3.8 3.9 4 4.1 4.2 4.3
##   AF     0 0   0   1   1   2   1   1 0   1   0   0   0   0 2   1   1   1
##   AS     1 2   4   0   2   1   0   1 0   0   0   0   0   0 0   0   0   0
##   EU     0 0   0   0   0   0   0   0 0   0   0   0   1   1 0   0   0   0
##   NAM    0 0   2   0   0   0   0   0 1   0   1   1   0   0 0   0   1   0
##   OC     0 1   0   1   0   0   1   0 0   0   0   0   0   0 0   0   0   0
##   SA     0 0   0   0   0   0   0   0 0   0   0   0   1   0 0   0   2   0
##      
##        4.4 4.6 4.7 4.9 5 5.4 5.5 5.6 5.7 5.8 5.9 6.1 6.2 6.3 6.4 6.5 6.6
##   AF     0   0   2   0 0   1   0   0   1   2   1   0   0   1   0   0   0
##   AS     0   1   0   1 1   0   0   0   0   0   0   0   1   0   1   0   0
##   EU     0   1   0   2 0   1   0   0   0   0   0   0   0   1   0   1   2
##   NAM    1   0   0   1 0   0   1   0   0   0   1   0   1   3   1   0   1
##   OC     0   0   0   0 0   0   0   0   0   0   1   0   0   0   0   0   0
##   SA     0   0   0   0 0   0   0   1   0   0   0   1   0   0   0   0   1
##      
##        6.7 6.8 6.9 7 7.1 7.2 7.3 7.6 7.7 8.2 8.3 8.7 8.9 9.1 9.3 9.4 9.5
##   AF     1   2   0 0   0   0   0   0   0   1   1   0   1   1   0   0   0
##   AS     0   1   0 1   0   0   0   0   0   0   0   0   0   0   0   0   0
##   EU     1   0   0 0   0   1   0   0   0   1   1   0   1   0   0   1   1
##   NAM    0   1   0 0   0   1   0   0   1   1   0   1   0   0   0   0   0
##   OC     0   0   1 1   0   0   0   0   0   0   0   0   0   0   1   0   0
##   SA     0   0   0 0   1   1   1   1   1   0   1   0   0   0   0   0   0
##      
##        9.6 9.7 9.8 10 10.1 10.2 10.3 10.4 10.5 10.6 10.9 11 11.3 11.4 11.5
##   AF     0   0   0  0    0    0    0    0    0    0    0  0    0    0    0
##   AS     0   0   1  0    0    0    0    0    0    0    0  0    0    0    1
##   EU     1   1   0  2    0    2    1    3    2    1    1  1    2    3    0
##   NAM    0   0   0  0    1    0    0    0    0    0    0  0    0    0    0
##   OC     0   0   0  0    0    0    0    1    0    0    0  0    0    0    0
##   SA     0   0   0  0    0    0    0    0    0    0    0  0    0    0    0
##      
##        11.8 11.9 12.4 12.9 14.4
##   AF      0    0    0    0    0
##   AS      0    0    0    0    0
##   EU      2    0    1    1    1
##   NAM     0    1    0    0    0
##   OC      0    0    0    0    0
##   SA      0    0    0    0    0
```

Hypothesis Testing Cont.

```
chi_test$expected
```

```
##
##              0       0.1       0.2       0.3       0.4       0.5
##   AF  3.5699482 1.9222798 1.0984456 0.8238342 0.5492228 0.8238342
##   AS  2.9637306 1.5958549 0.9119171 0.6839378 0.4559585 0.6839378
##   EU  3.0310881 1.6321244 0.9326425 0.6994819 0.4663212 0.6994819
##   NAM 1.5492228 0.8341969 0.4766839 0.3575130 0.2383420 0.3575130
##   OC  1.0777202 0.5803109 0.3316062 0.2487047 0.1658031 0.2487047
##   SA  0.8082902 0.4352332 0.2487047 0.1865285 0.1243523 0.1865285
##
##            0.6       0.7        0.8       0.9         1       1.1
##   AF  0.5492228 0.8238342 0.27461140 0.5492228 1.0984456 0.8238342
##   AS  0.4559585 0.6839378 0.22797927 0.4559585 0.9119171 0.6839378
##   EU  0.4663212 0.6994819 0.23316062 0.4663212 0.9326425 0.6994819
##   NAM 0.2383420 0.3575130 0.11917098 0.2383420 0.4766839 0.3575130
##   OC  0.1658031 0.2487047 0.08290155 0.1658031 0.3316062 0.2487047
##   SA  0.1243523 0.1865285 0.06217617 0.1243523 0.2487047 0.1865285
##
##             1.2       1.3        1.4       1.5       1.7       1.8
##   AF  0.27461140 1.0984456 0.27461140 0.8238342 0.5492228 0.8238342
##   AS  0.22797927 0.9119171 0.22797927 0.6839378 0.4559585 0.6839378
##   EU  0.23316062 0.9326425 0.23316062 0.6994819 0.4663212 0.6994819
##   NAM 0.11917098 0.4766839 0.11917098 0.3575130 0.2383420 0.3575130
##   OC  0.08290155 0.3316062 0.08290155 0.2487047 0.1658031 0.2487047
##   SA  0.06217617 0.2487047 0.06217617 0.1865285 0.1243523 0.1865285
##
##             1.9         2       2.2       2.3       2.4       2.5
##   AF  0.27461140 0.8238342 1.6476684 0.5492228 0.8238342 0.8238342
##   AS  0.22797927 0.6839378 1.3678756 0.4559585 0.6839378 0.6839378
##   EU  0.23316062 0.6994819 1.3989637 0.4663212 0.6994819 0.6994819
##   NAM 0.11917098 0.3575130 0.7150259 0.2383420 0.3575130 0.3575130
##   OC  0.08290155 0.2487047 0.4974093 0.1658031 0.2487047 0.2487047
##   SA  0.06217617 0.1865285 0.3730570 0.1243523 0.1865285 0.1865285
##
##            2.6       2.8          3        3.1        3.4        3.5
##   AF  0.5492228 0.5492228 0.27461140 0.27461140 0.27461140 0.27461140
##   AS  0.4559585 0.4559585 0.22797927 0.22797927 0.22797927 0.22797927
##   EU  0.4663212 0.4663212 0.23316062 0.23316062 0.23316062 0.23316062
##   NAM 0.2383420 0.2383420 0.11917098 0.11917098 0.11917098 0.11917098
##   OC  0.1658031 0.1658031 0.08290155 0.08290155 0.08290155 0.08290155
##   SA  0.1243523 0.1243523 0.06217617 0.06217617 0.06217617 0.06217617
##
##            3.8        3.9         4        4.1       4.2        4.3
##   AF  0.5492228 0.27461140 0.5492228 0.27461140 1.0984456 0.27461140
##   AS  0.4559585 0.22797927 0.4559585 0.22797927 0.9119171 0.22797927
##   EU  0.4663212 0.23316062 0.4663212 0.23316062 0.9326425 0.23316062
##   NAM 0.2383420 0.11917098 0.2383420 0.11917098 0.4766839 0.11917098
##   OC  0.1658031 0.08290155 0.1658031 0.08290155 0.3316062 0.08290155
##   SA  0.1243523 0.06217617 0.1243523 0.06217617 0.2487047 0.06217617
##
##             4.4       4.6       4.7       4.9          5       5.4
##   AF  0.27461140 0.5492228 0.5492228 1.0984456 0.27461140 0.5492228
##   AS  0.22797927 0.4559585 0.4559585 0.9119171 0.22797927 0.4559585
##   EU  0.23316062 0.4663212 0.4663212 0.9326425 0.23316062 0.4663212
##   NAM 0.11917098 0.2383420 0.2383420 0.4766839 0.11917098 0.2383420
##   OC  0.08290155 0.1658031 0.1658031 0.3316062 0.08290155 0.1658031
##   SA  0.06217617 0.1243523 0.1243523 0.2487047 0.06217617 0.1243523
##
##             5.5        5.6        5.7       5.8       5.9        6.1
##   AF  0.27461140 0.27461140 0.27461140 0.5492228 0.8238342 0.27461140
##   AS  0.22797927 0.22797927 0.22797927 0.4559585 0.6839378 0.22797927
##   EU  0.23316062 0.23316062 0.23316062 0.4663212 0.6994819 0.23316062
##   NAM 0.11917098 0.11917098 0.11917098 0.2383420 0.3575130 0.11917098
##   OC  0.08290155 0.08290155 0.08290155 0.1658031 0.2487047 0.08290155
##   SA  0.06217617 0.06217617 0.06217617 0.1243523 0.1865285 0.06217617
##
##            6.2       6.3       6.4        6.5       6.6       6.7
##   AF  0.5492228 1.3730570 0.5492228 0.27461140 1.0984456 0.5492228
##   AS  0.4559585 1.1398964 0.4559585 0.22797927 0.9119171 0.4559585
##   EU  0.4663212 1.1658031 0.4663212 0.23316062 0.9326425 0.4663212
##   NAM 0.2383420 0.5958549 0.2383420 0.11917098 0.4766839 0.2383420
##   OC  0.1658031 0.4145078 0.1658031 0.08290155 0.3316062 0.1658031
##   SA  0.1243523 0.3108808 0.1243523 0.06217617 0.2487047 0.1243523
```

```
##
##             6.8        6.9         7        7.1        7.2        7.3
##    AF  1.0984456 0.27461140 0.5492228 0.27461140 0.8238342 0.27461140
##    AS  0.9119171 0.22797927 0.4559585 0.22797927 0.6839378 0.22797927
##    EU  0.9326425 0.23316062 0.4663212 0.23316062 0.6994819 0.23316062
##    NAM 0.4766839 0.11917098 0.2383420 0.11917098 0.3575130 0.11917098
##    OC  0.3316062 0.08290155 0.1658031 0.08290155 0.2487047 0.08290155
##    SA  0.2487047 0.06217617 0.1243523 0.06217617 0.1865285 0.06217617
##
##              7.6       7.7       8.2       8.3        8.7       8.9
##    AF  0.27461140 0.5492228 0.8238342 0.8238342 0.27461140 0.5492228
##    AS  0.22797927 0.4559585 0.6839378 0.6839378 0.22797927 0.4559585
##    EU  0.23316062 0.4663212 0.6994819 0.6994819 0.23316062 0.4663212
##    NAM 0.11917098 0.2383420 0.3575130 0.3575130 0.11917098 0.2383420
##    OC  0.08290155 0.1658031 0.2487047 0.2487047 0.08290155 0.1658031
##    SA  0.06217617 0.1243523 0.1865285 0.1865285 0.06217617 0.1243523
##
##              9.1        9.3        9.4        9.5        9.6        9.7
##    AF  0.27461140 0.27461140 0.27461140 0.27461140 0.27461140 0.27461140
##    AS  0.22797927 0.22797927 0.22797927 0.22797927 0.22797927 0.22797927
##    EU  0.23316062 0.23316062 0.23316062 0.23316062 0.23316062 0.23316062
##    NAM 0.11917098 0.11917098 0.11917098 0.11917098 0.11917098 0.11917098
##    OC  0.08290155 0.08290155 0.08290155 0.08290155 0.08290155 0.08290155
##    SA  0.06217617 0.06217617 0.06217617 0.06217617 0.06217617 0.06217617
##
##              9.8        10       10.1      10.2       10.3      10.4
##    AF  0.27461140 0.5492228 0.27461140 0.5492228 0.27461140 1.0984456
##    AS  0.22797927 0.4559585 0.22797927 0.4559585 0.22797927 0.9119171
##    EU  0.23316062 0.4663212 0.23316062 0.4663212 0.23316062 0.9326425
##    NAM 0.11917098 0.2383420 0.11917098 0.2383420 0.11917098 0.4766839
##    OC  0.08290155 0.1658031 0.08290155 0.1658031 0.08290155 0.3316062
##    SA  0.06217617 0.1243523 0.06217617 0.1243523 0.06217617 0.2487047
##
##            10.5       10.6       10.9         11      11.3      11.4
##    AF  0.5492228 0.27461140 0.27461140 0.27461140 0.5492228 0.8238342
##    AS  0.4559585 0.22797927 0.22797927 0.22797927 0.4559585 0.6839378
##    EU  0.4663212 0.23316062 0.23316062 0.23316062 0.4663212 0.6994819
##    NAM 0.2383420 0.11917098 0.11917098 0.11917098 0.2383420 0.3575130
##    OC  0.1658031 0.08290155 0.08290155 0.08290155 0.1658031 0.2487047
##    SA  0.1243523 0.06217617 0.06217617 0.06217617 0.1243523 0.1865285
##
##             11.5      11.8       11.9       12.4       12.9       14.4
##    AF  0.27461140 0.5492228 0.27461140 0.27461140 0.27461140 0.27461140
##    AS  0.22797927 0.4559585 0.22797927 0.22797927 0.22797927 0.22797927
##    EU  0.23316062 0.4663212 0.23316062 0.23316062 0.23316062 0.23316062
##    NAM 0.11917098 0.2383420 0.11917098 0.11917098 0.11917098 0.11917098
##    OC  0.08290155 0.1658031 0.08290155 0.08290155 0.08290155 0.08290155
##    SA  0.06217617 0.1243523 0.06217617 0.06217617 0.06217617 0.06217617
```

Hypthesis Testing Cont.

- To test H0, the Chi-square statistic is calculated as follows:

$$X2 = ??(Oij - Eij)^2/Eij$$

Hypthesis Testing Cont.

- x2 = 546.34
- p-value < 0.01
- Decision: Reject H0
- Conclusion: The result shows that there is statistically significant association between the total liters of pure alcohol and continent.

Discussion

- The results of this investigation demonstrate that total_litres_of_pure_alcohol and continents as a statistically significant association.
- Strength:
- This dataset was easy to observe, investigate numerous variables and perform the required hypothesis testing.
- Limitations:
- This investigation is constrained to limited types of alcoholic beverages (beer, spirit, wine).
- It contains the countries that don't serve alcohol.
- Many countries import and export alcoholic beverages so, the correct volume of pure alcohol for a particular continent is difficult to obtain.

Discussion Cont.

- Future investigation:
- Besides, an investigation in the future should have an expansion in the dataset where it includes all types of alcoholic beverages.
- From the findings of the above investigation conducted, a Chi-square Test of Association found that there is statistically significant association between the pureness of alcohol across various continents. Thus, the investigation concludes the Europe continent serves more pure of alcohol.

References

- Science X 2018, photograph, viewed on 5 October 2018 https://3c1703fe8d.site.internapcdn.net/newman/gfx/news/hires/2017/ismixingdrin.jpg
- Navneeth 2017, Drinks CSV, Drinks quality dataset, Viewed on 6 October 2018 https://www.kaggle.com/navneethc/drinks