

Statistical Analysis

US Consumer Complaints dataset contains 18 variables. Most of the variables like company, product, sub_product, issue and sub_issue are considered as categorical variables

Chi-square Tests

Chi-Square tests assumes a null hypothesis and an alternate hypothesis. If the p-value is less than the pre-determined value (0.05) then null hypothesis will be rejected. The conclusion is those two categorical values are dependent.

Reading the data

```
complaint <- read_csv("consumer_complaints.csv")

## Parsed with column specification:
## cols(
##   date_received = col_character(),
##   product = col_character(),
##   sub_product = col_character(),
##   issue = col_character(),
##   sub_issue = col_character(),
##   consumer_complaint_narrative = col_character(),
##   company_public_response = col_character(),
##   company = col_character(),
##   state = col_character(),
##   zipcode = col_character(),
##   tags = col_character(),
##   consumer_consent_provided = col_character(),
##   submitted_via = col_character(),
##   date_sent_to_company = col_character(),
##   company_response_to_consumer = col_character(),
##   timely_response = col_character(),
##   `consumer_disputed?` = col_character(),
##   complaint_id = col_integer()
## )

comp_prod <- table(complaint$company, complaint$product)
complaint$product <- as.factor(complaint$product)
complaint$company <- as.factor(complaint$company)
complaint$sub_product <- as.factor(complaint$sub_product)
complaint$issue <- as.factor(complaint$issue)
complaint$submitted_via <- as.factor(complaint$submitted_via)
```

Chi-Square tests

Company and Product are dependent.

```
chisq.test(complaint$company, complaint$product, correct = FALSE)

## Warning in chisq.test(complaint$company, complaint$product, correct =
## FALSE): Chi-squared approximation may be incorrect
```

```
##
## Pearson's Chi-squared test
##
## data:  complaint$company and complaint$product
## X-squared = 2670800, df = 36040, p-value < 2.2e-16
chisq.test(complaint$issue,complaint$submitted_via,correct = FALSE)

## Warning in chisq.test(complaint$issue, complaint$submitted_via, correct =
## FALSE): Chi-squared approximation may be incorrect
##
## Pearson's Chi-squared test
##
## data:  complaint$issue and complaint$submitted_via
## X-squared = 110470, df = 470, p-value < 2.2e-16
From above results issue and subitted_via also dependent variables
comp_disp <- complaint %>%
  select(company,`consumer_disputed?`)%>%
  na.omit()

comp_disp_table <- table(comp_disp$company,comp_disp$`consumer_disputed?`)

chisq.test(comp_disp_table,correct = FALSE)

## Warning in chisq.test(comp_disp_table, correct = FALSE): Chi-squared
## approximation may be incorrect
##
## Pearson's Chi-squared test
##
## data:  comp_disp_table
## X-squared = 10723, df = 3604, p-value < 2.2e-16
chisq.test(complaint$company,complaint$`consumer_disputed?`,correct = FALSE)

## Warning in chisq.test(complaint$company, complaint$`consumer_disputed?`, :
## Chi-squared approximation may be incorrect
##
## Pearson's Chi-squared test
##
## data:  complaint$company and complaint$`consumer_disputed?`
## X-squared = 10723, df = 3604, p-value < 2.2e-16
chisq.test(complaint$timely_response,complaint$`consumer_disputed?`,correct = FALSE)

##
## Pearson's Chi-squared test
##
## data:  complaint$timely_response and complaint$`consumer_disputed?`
## X-squared = 569.37, df = 1, p-value < 2.2e-16
```

The variables for which chi-squared tests are considered are proven to be dependent.