Statistical Methods for Discrete Response, Time Series, and Panel Data: Live ession 4

Professor Jeffrey Yau

Main Topics Covered in Lecture 4:

- Multinomial probability distribution
- IJ contingency tables and inference using contingency tables
- The notion of independence
- Nominal response models
- Odds ratios in the context of nominal response models
- Ordinal logistical regression model
- Estimation and statistical inference of these models

Required Readings:

BL2015: Christopher R. Bilder and Thomas M. Loughin. Analysis of Categorical Data with R. CRC Press. 2015.

- Ch.3 (Skip Sections 3.4.3, 3.5)

In this exercise, we want to model voters' self identified party affiliation using their demographic characteristics and a handful of self-indentifying variables. The data was obtained from the **American National Election Survey**, which conducted a survey several months prior to the 2016 American Presidential elections. *Note that the original survey data uses survey weights, which we will not use here.*

The dataset "*voters.csv*" contains a handful of variables from the survey, and these variables have been cleaned and modified for this exercise. This dataset contains the following variables:

Variable Name	Explanations
party	Categorical variable indicating respondents' party affiliation: Democrat, Independent, Republican
Presjob	A seven point scale indicating respondents' evaluation of President Obama. 1 = Very strongly approve; 7 = Very strongly disapprove
Srv_spend	Seven point scale representing the degree to which respondents believe that the government should provide or should not provide services: 1 = Government should provide many fewer services; 7 = Government should provide many more services.
age	Respondents' age, as of 2016.
race_white	Dummy variable taking a value of one if the respondent is white and is zero otherwise.
female	Dummy variable taking a value of one if the respondent is female and is zero otherwise.

#EDA

Setup Codes and Load Data

```
knitr::opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)

# Load Libraries
library(car)
library(Hmisc)
library(plyr)
library(dplyr)
library(skimr)
library(ggplot2)
library(stargazer)

library(gmodels) # For cross tabulation (SAS and SPSS style)
library(MASS)
library(mcprofile)
library(vcd)
library(nnet)

df <- read.csv("live_session_04_data.csv", stringsAsFactors = FALSE, header = TRUE, sep = ",")

# Make data
```

```
str(df)
```

```
## 'data.frame':    1200 obs. of  11 variables:
##  $ ftwhite   : int  100 74 50 64 58 51 70 70 50 90 ...
##  $ ftblack   : int  100 6 50 61 61 50 100 70 50 75 ...
##  $ ftmuslim  : int  20 22 5 61 22 11 100 40 12 72 ...
##  $ presjob   : int  1 3 7 2 7 7 2 7 7 2 ...
##  $ srv_spend : int  7 6 2 6 1 1 7 3 1 6 ...
##  $ crimespend: int  5 2 5 4 4 7 2 4 4 3 ...
##  $ party     : chr  "Democrat" "Independent" "Republican" "Democrat" ...
##  $ ideo5     : int  NA 2 4 2 4 4 1 5 4 2 ...
##  $ age       : int  56 59 53 36 42 58 38 65 43 80 ...
##  $ race_white: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ female    : int  0 1 0 0 0 0 0 0 0 0 ...
```

```
voters <- df %>%
  dplyr::select(party, presjob, srv_spend,
         age, female, race_white)
voters$presjob <- revalue(as.factor(voters$presjob),
                          c("1"="Approve", "2"="Approve",
                            "3"="Neutral", "4"="Neutral",
                            "5"="Neutral", "6"="Not Approve",
                            "7"="Not Approve"))
voters$srv_spend <- revalue(as.factor(voters$srv_spend),
                          c("1"="Low", "2"="Low", "3"="Low",
                            "4"="Medium", "5"="Medium",
                            "997"="Medium", "6"="High", "7"="High"))
voters$female <- revalue(as.factor(voters$female),
                         c("0"="Male", "1"="Female"))
voters$race_white <- revalue(as.factor(voters$race_white),
                             c("0"="Non-White", "1"="White"))

str(voters)
```

```
## 'data.frame':    1200 obs. of  6 variables:
##  $ party     : chr  "Democrat" "Independent" "Republican" "Democrat" ...
##  $ presjob   : Factor w/ 3 levels "Approve","Neutral",..: 1 2 3 1 3 3 1 3 3 1 ...
##  $ srv_spend : Factor w/ 3 levels "Low","Medium",..: 3 3 1 3 1 1 3 1 1 3 ...
##  $ age       : int  56 59 53 36 42 58 38 65 43 80 ...
##  $ female    : Factor w/ 2 levels "Male","Female": 1 2 1 1 1 1 1 1 1 1 ...
##  $ race_white: Factor w/ 2 levels "Non-White","White": 2 2 2 2 2 2 2 2 2 2 ...
```

```
skim(voters)
```

```
## Skim summary statistics
##  n obs: 1200
##  n variables: 6
##
## -- Variable type:character -----------------------
##  variable missing complete    n min max empty n_unique
##     party      81     1119 1200   8  11     0        3
##
## -- Variable type:factor --------------------------
##    variable missing complete    n n_unique
##      female       0     1200 1200        2
```

```
##     presjob       0     1200 1200        3
##  race_white       0     1200 1200        2
##   srv_spend       0     1200 1200        3
##                           top_counts ordered
##            Fem: 630, Mal: 570, NA: 0   FALSE
##  Not: 492, App: 453, Neu: 255, NA: 0   FALSE
##            Whi: 875, Non: 325, NA: 0   FALSE
##  Med: 491, Low: 406, Hig: 303, NA: 0   FALSE
##
## -- Variable type:integer -------------------------
##  variable missing complete    n  mean    sd p0 p25 p50   p75 p100     hist
##       age       0     1200 1200 48.06 16.99 19  34  48 61.25   95
```

```
write.csv(voters, file = "voters.csv", sep = ",",
          row.names = FALSE, col.names = TRUE)

voters <- read.csv("voters.csv", stringsAsFactors = FALSE, header = TRUE, sep = ",")

# Convert all the character variables to factor variables
voters <- voters%>%
  dplyr::mutate_if(sapply(voters, is.character), as.factor)
```

Breakout-room Discussion: - Discuss the structure of the data - Discuss missing values and how you would typically handle them at work - Discuss the patterns of these variables - Add additional tables and plots to enhance your EDA where needed

```
str(voters)
```

```
## 'data.frame':    1200 obs. of  6 variables:
##  $ party     : Factor w/ 3 levels "Democrat","Independent",..: 1 2 3 1 NA 2 1 3 2 1 ...
##  $ presjob   : Factor w/ 3 levels "Approve","Neutral",..: 1 2 3 1 3 3 1 3 3 1 ...
##  $ srv_spend : Factor w/ 3 levels "High","Low","Medium": 1 1 2 1 2 2 1 2 2 1 ...
##  $ age       : int  56 59 53 36 42 58 38 65 43 80 ...
##  $ female    : Factor w/ 2 levels "Female","Male": 2 1 2 2 2 2 2 2 2 2 ...
##  $ race_white: Factor w/ 2 levels "Non-White","White": 2 2 2 2 2 2 2 2 2 2 ...
```

```
skim(voters)
```

```
## Skim summary statistics
##  n obs: 1200
##  n variables: 6
##
## -- Variable type:factor -------------------------
##    variable missing complete    n n_unique
##      female       0     1200 1200        2
##       party      81     1119 1200        3
##     presjob       0     1200 1200        3
##  race_white       0     1200 1200        2
##   srv_spend       0     1200 1200        3
##                           top_counts ordered
##             Fem: 630, Mal: 570, NA: 0   FALSE
##  Dem: 459, Ind: 380, Rep: 280, NA: 81   FALSE
##   Not: 492, App: 453, Neu: 255, NA: 0   FALSE
##             Whi: 875, Non: 325, NA: 0   FALSE
##   Med: 491, Low: 406, Hig: 303, NA: 0   FALSE
##
```

```
## -- Variable type:integer --------------------------
##  variable missing complete    n  mean    sd p0 p25 p50   p75 p100     hist
##       age       0     1200 1200 48.06 16.99 19  34  48 61.25   95
```

```
describe(voters)
```

```
## voters
##
##  6  Variables      1200  Observations
## ---------------------------------------------------------------------------
## party
##        n  missing distinct
##     1119       81        3
##
## Value         Democrat Independent  Republican
## Frequency          459         380         280
## Proportion        0.41        0.34        0.25
## ---------------------------------------------------------------------------
## presjob
##        n  missing distinct
##     1200        0        3
##
## Value          Approve     Neutral Not Approve
## Frequency          453         255         492
## Proportion       0.378       0.212       0.410
## ---------------------------------------------------------------------------
## srv_spend
##        n  missing distinct
##     1200        0        3
##
## Value        High    Low Medium
## Frequency     303    406    491
## Proportion  0.252  0.338  0.409
## ---------------------------------------------------------------------------
## age
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1200        0       73        1    48.06    19.53    22.00    25.00
##      .25      .50      .75      .90      .95
##    34.00    48.00    61.25    70.00    76.00
##
## lowest : 19 20 21 22 23, highest: 89 90 91 92 95
## ---------------------------------------------------------------------------
## female
##        n  missing distinct
##     1200        0        2
##
## Value      Female   Male
## Frequency     630    570
## Proportion  0.525  0.475
## ---------------------------------------------------------------------------
## race_white
##        n  missing distinct
##     1200        0        2
##
## Value      Non-White     White
```

```
## Frequency        325       875
## Proportion     0.271     0.729
## ------------------------------------------------------------------------
```

```
# voters[!complete.cases(voters),]
sapply(voters, function(x) sum(is.na(x)))
```

```
##      party    presjob  srv_spend        age     female race_white
##         81          0          0          0          0          0
```

```
# Keep only the complete cases in the dataset
voters2 <- voters[complete.cases(voters), ]

# Reorder the categories of srv_spend
voters2$srv_spend <- ordered(voters2$srv_spend, levels = c("Low",
    "Medium", "High"))

# Attach the dataste
attach(voters2)
```

Pause and Discuss: Missing values For now, we would simply exclude them in our analysis. *In practice, you do not just want to throw away observations without any investigation.*

EDA:

```
# Descriptive statistics
str(voters2)
```

```
## 'data.frame':    1119 obs. of  6 variables:
##  $ party     : Factor w/ 3 levels "Democrat","Independent",..: 1 2 3 1 2 1 3 2 1 3 ...
##  $ presjob   : Factor w/ 3 levels "Approve","Neutral",..: 1 2 3 1 3 1 3 3 1 3 ...
##  $ srv_spend : Ord.factor w/ 3 levels "Low"<"Medium"<..: 3 3 1 3 1 3 1 1 3 1 ...
##  $ age       : int  56 59 53 36 58 38 65 43 80 38 ...
##  $ female    : Factor w/ 2 levels "Female","Male": 2 1 2 2 2 2 2 2 2 2 ...
##  $ race_white: Factor w/ 2 levels "Non-White","White": 2 2 2 2 2 2 2 2 2 2 ...
```

```
skim(voters2)
```

```
## Skim summary statistics
##  n obs: 1119
##  n variables: 6
##
## -- Variable type:factor -------------------------
##    variable missing complete    n n_unique
##      female       0     1119 1119        2
##       party       0     1119 1119        3
##     presjob       0     1119 1119        3
##  race_white       0     1119 1119        2
##   srv_spend       0     1119 1119        3
##                            top_counts ordered
##            Fem: 593, Mal: 526, NA: 0   FALSE
##  Dem: 459, Ind: 380, Rep: 280, NA: 0   FALSE
##  Not: 446, App: 439, Neu: 234, NA: 0   FALSE
##            Whi: 813, Non: 306, NA: 0   FALSE
##  Med: 458, Low: 369, Hig: 292, NA: 0    TRUE
```

```
##
## -- Variable type:integer -------------------------
##  variable missing complete    n  mean    sd p0 p25 p50 p75 p100     hist
##       age       0     1119 1119 48.25 17.01 19  34  49  62   95
```

```
describe(voters2)
```

```
## voters2
##
##  6  Variables      1119  Observations
## ---------------------------------------------------------------------------
## party
##        n  missing distinct
##     1119        0        3
##
## Value         Democrat Independent  Republican
## Frequency          459         380         280
## Proportion        0.41        0.34        0.25
## ---------------------------------------------------------------------------
## presjob
##        n  missing distinct
##     1119        0        3
##
## Value          Approve     Neutral Not Approve
## Frequency          439         234         446
## Proportion       0.392       0.209       0.399
## ---------------------------------------------------------------------------
## srv_spend
##        n  missing distinct
##     1119        0        3
##
## Value         Low Medium   High
## Frequency     369    458    292
## Proportion  0.330  0.409  0.261
## ---------------------------------------------------------------------------
## age
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1119        0       72        1    48.25    19.56       22       25
##      .25      .50      .75      .90      .95
##       34       49       62       71       76
##
## lowest : 19 20 21 22 23, highest: 89 90 91 92 95
## ---------------------------------------------------------------------------
## female
##        n  missing distinct
##     1119        0        2
##
## Value      Female   Male
## Frequency     593    526
## Proportion   0.53   0.47
## ---------------------------------------------------------------------------
## race_white
##        n  missing distinct
##     1119        0        2
##
```

```
## Value      Non-White     White
## Frequency        306       813
## Proportion     0.273     0.727
## ------------------------------------------------------------------------
```

```
# Univariate Analysis
apply(voters2, 2, table)
```

```
## $party
##
##    Democrat Independent  Republican
##         459         380         280
##
## $presjob
##
##     Approve     Neutral Not Approve
##         439         234         446
##
## $srv_spend
##
##   High    Low Medium
##    292    369    458
##
## $age
##
## 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
## 12 16 13 19 17 21 19 23 27 20 19 20 15 14 19 19 19 24 16 25 22 20 23 17 30
## 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68
## 21 10 12  8 17 13  9 15 20 16 21 29 26 22 20 24 33 30 36 26 20 20 18 16  9
## 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 89 90 91 92 95
## 15 10 15  8 14 10  8 11  6 10  6  8  1  2  4  3  2  1  1  2  1  1
##
## $female
##
## Female   Male
##    593    526
##
## $race_white
##
## Non-White     White
##       306       813
```

```
exam_cat_var = function(var.names) {
    round(prop.table(table(var.names)), 2)
}
apply(voters2, 2, exam_cat_var)
```

```
## $party
## var.names
##    Democrat Independent  Republican
##        0.41        0.34        0.25
##
## $presjob
## var.names
##     Approve     Neutral Not Approve
```

```
##        0.39        0.21        0.40
##
## $srv_spend
## var.names
##   High    Low Medium
##   0.26   0.33   0.41
##
## $age
## var.names
##   19   20   21   22   23   24   25   26   27   28   29   30   31   32   33
## 0.01 0.01 0.01 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.01 0.01 0.02
##   34   35   36   37   38   39   40   41   42   43   44   45   46   47   48
## 0.02 0.02 0.02 0.01 0.02 0.02 0.02 0.02 0.02 0.03 0.02 0.01 0.01 0.01 0.02
##   49   50   51   52   53   54   55   56   57   58   59   60   61   62   63
## 0.01 0.01 0.01 0.02 0.01 0.02 0.03 0.02 0.02 0.02 0.02 0.03 0.03 0.03 0.02
##   64   65   66   67   68   69   70   71   72   73   74   75   76   77   78
## 0.02 0.02 0.02 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01
##   79   80   81   82   83   84   85   89   90   91   92   95
## 0.01 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
##
## $female
## var.names
## Female   Male
##   0.53   0.47
##
## $race_white
## var.names
## Non-White     White
##      0.27      0.73
```

```r
# Bivariate Analysis
cross_tab = function(xvar, yvar) {
    CrossTable(xvar, yvar, digits = 2, prop.c = FALSE, prop.t = FALSE,
        chisq = TRUE)
}
# President Approval by Party
cross_tab(voters2$presjob, voters2$party)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |-------------------------|
##
##
## Total Observations in Table:  1119
##
##
##              | yvar
##         xvar |    Democrat | Independent |  Republican |   Row Total |
## -------------|-------------|-------------|-------------|-------------|
##      Approve |         331 |          88 |          20 |         439 |
```

```
##              |      126.50 |       25.02 |       73.49 |              |
##              |        0.75 |        0.20 |        0.05 |         0.39 |
## -------------|-------------|-------------|-------------|--------------|
##      Neutral |         100 |          99 |          35 |          234 |
##              |        0.17 |        4.80 |        9.47 |              |
##              |        0.43 |        0.42 |        0.15 |         0.21 |
## -------------|-------------|-------------|-------------|--------------|
##  Not Approve |          28 |         193 |         225 |          446 |
##              |      131.23 |       11.40 |      115.23 |              |
##              |        0.06 |        0.43 |        0.50 |         0.40 |
## -------------|-------------|-------------|-------------|--------------|
## Column Total |         459 |         380 |         280 |         1119 |
## -------------|-------------|-------------|-------------|--------------|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  497.3134     d.f. =  4     p =  2.553304e-106
##
##
##
```

```
# Spending Sentiment by Party
cross_tab(voters2$srv_spend, voters2$party)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |-------------------------|
##
##
## Total Observations in Table:  1119
##
##
##              | yvar
##         xvar |    Democrat | Independent |  Republican |   Row Total |
## -------------|-------------|-------------|-------------|-------------|
##          Low |          48 |         158 |         163 |         369 |
##              |       70.58 |        8.53 |       54.09 |             |
##              |        0.13 |        0.43 |        0.44 |        0.33 |
## -------------|-------------|-------------|-------------|-------------|
##       Medium |         217 |         150 |          91 |         458 |
##              |        4.52 |        0.20 |        4.86 |             |
##              |        0.47 |        0.33 |        0.20 |        0.41 |
## -------------|-------------|-------------|-------------|-------------|
##         High |         194 |          72 |          26 |         292 |
##              |       46.00 |        7.44 |       30.32 |             |
##              |        0.66 |        0.25 |        0.09 |        0.26 |
```

```
## -------------|-------------|-------------|-------------|-------------|
## Column Total |         459 |         380 |         280 |        1119 |
## -------------|-------------|-------------|-------------|-------------|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  226.5263     d.f. =  4     p =  7.384957e-48
##
##
##
```

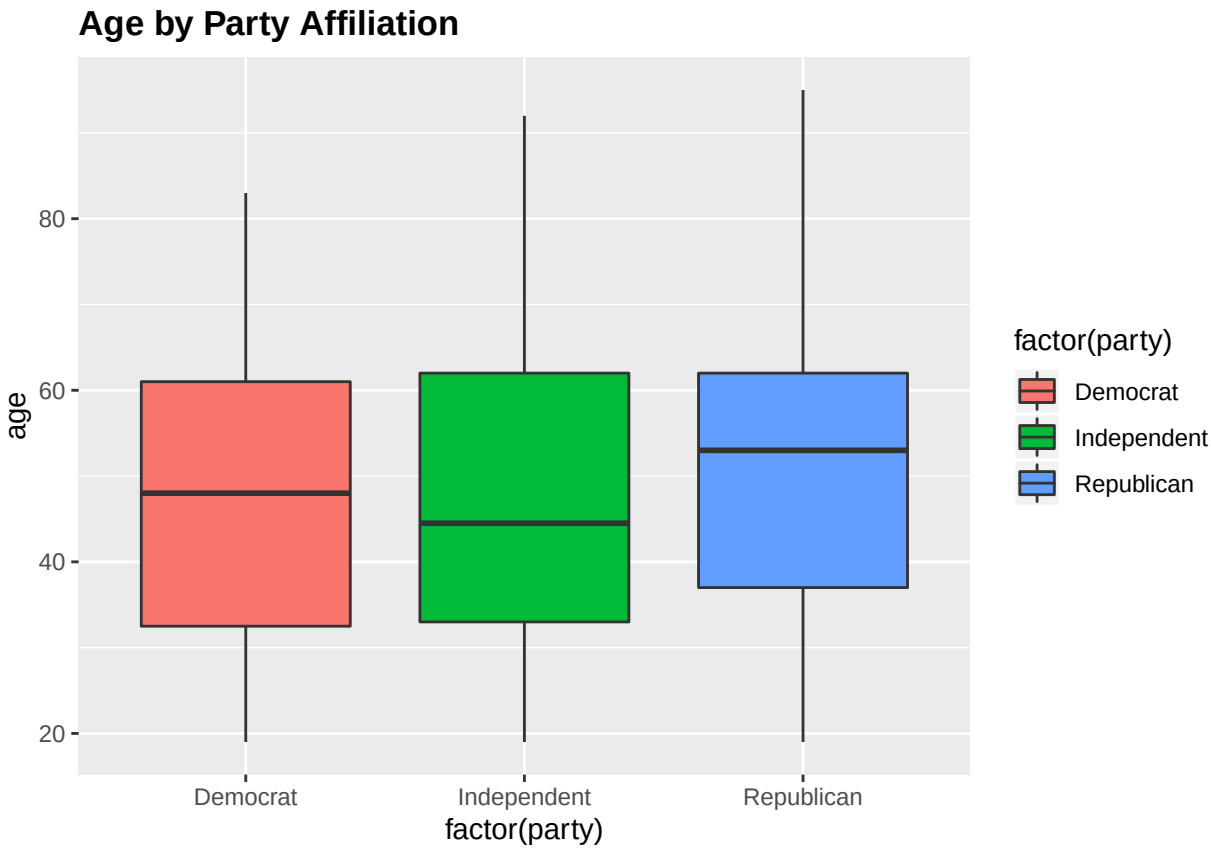
```
# Gender by Party
cross_tab(voters2$female, voters2$party)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |-------------------------|
##
##
## Total Observations in Table:  1119
##
##
##              | yvar
##         xvar |    Democrat | Independent |  Republican |   Row Total |
## -------------|-------------|-------------|-------------|-------------|
##       Female |         270 |         172 |         151 |         593 |
##              |        2.94 |        4.29 |        0.05 |             |
##              |        0.46 |        0.29 |        0.25 |        0.53 |
## -------------|-------------|-------------|-------------|-------------|
##         Male |         189 |         208 |         129 |         526 |
##              |        3.32 |        4.83 |        0.05 |             |
##              |        0.36 |        0.40 |        0.25 |        0.47 |
## -------------|-------------|-------------|-------------|-------------|
## Column Total |         459 |         380 |         280 |        1119 |
## -------------|-------------|-------------|-------------|-------------|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  15.47708     d.f. =  2     p =  0.0004357065
##
##
##
```

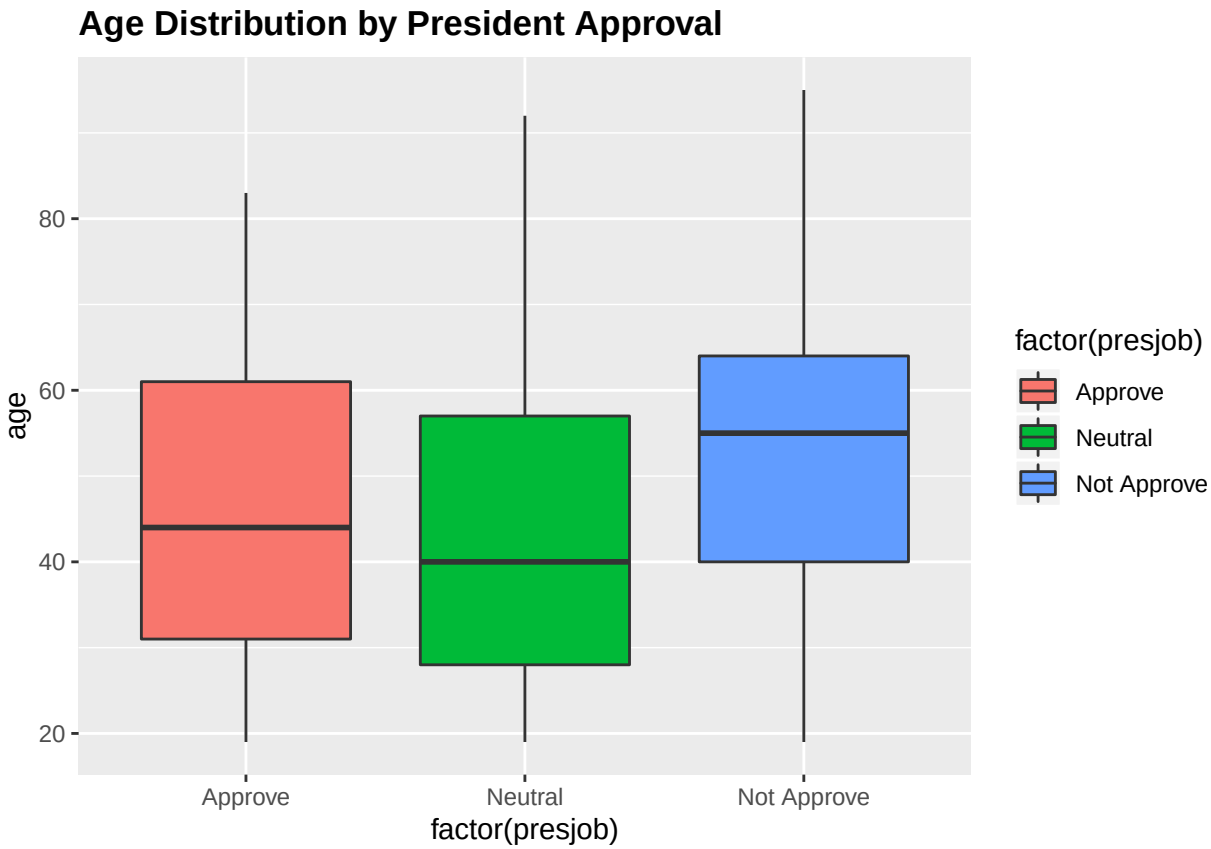
```r
# Race by Party
cross_tab(voters2$race_white, voters2$party)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |-------------------------|
##
##
## Total Observations in Table:  1119
##
##
##              | yvar
##         xvar |    Democrat | Independent |  Republican |   Row Total |
## -------------|-------------|-------------|-------------|-------------|
##    Non-White |         187 |          82 |          37 |         306 |
##              |       30.12 |        4.62 |       20.45 |             |
##              |        0.61 |        0.27 |        0.12 |        0.27 |
## -------------|-------------|-------------|-------------|-------------|
##        White |         272 |         298 |         243 |         813 |
##              |       11.34 |        1.74 |        7.70 |             |
##              |        0.33 |        0.37 |        0.30 |        0.73 |
## -------------|-------------|-------------|-------------|-------------|
## Column Total |         459 |         380 |         280 |        1119 |
## -------------|-------------|-------------|-------------|-------------|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  75.95634     d.f. =  2     p =  3.208416e-17
##
##
##
```

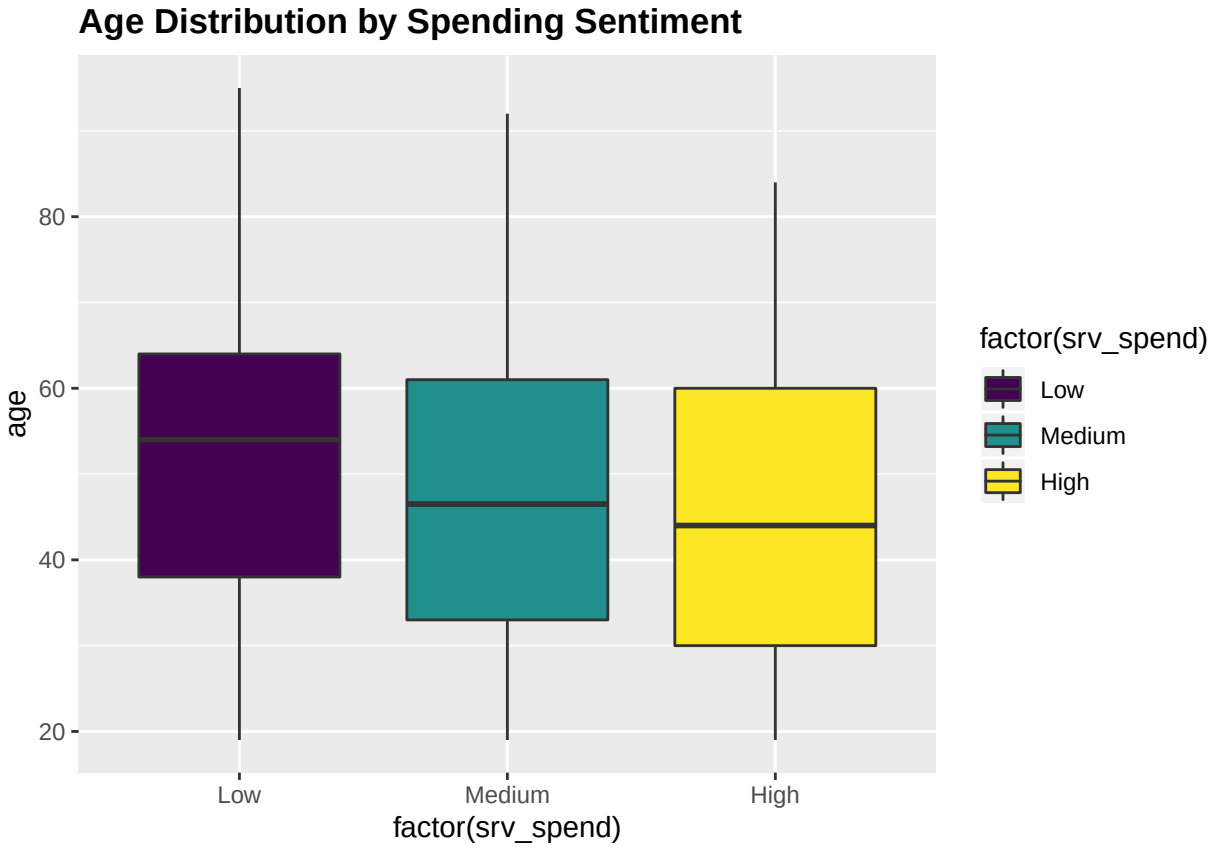
```r
# Age Distribution by Party
ggplot(voters2, aes(factor(party), age)) + geom_boxplot(aes(fill = factor(party))) +
    ggtitle("Age by Party Affiliation") + theme(plot.title = element_text(lineheight = 1,
    face = "bold"))
```




```
# Age Distribution by President Approval
ggplot(voters2, aes(factor(presjob), age)) + geom_boxplot(aes(fill = factor(presjob))) +
    ggtitle("Age Distribution by President Approval") + theme(plot.title = element_text(lineheight = 1,
    face = "bold"))
```




```
# Age Distribution by Spending Sentiment
ggplot(voters2, aes(factor(srv_spend), age)) + geom_boxplot(aes(fill = factor(srv_spend))) +
    ggtitle("Age Distribution by Spending Sentiment") + theme(plot.title = element_text(lineheight = 1,
    face = "bold"))
```




```
# President Approval by Spending Sentiment, Gender, and Race
cross_tab(voters2$srv_spend, voters2$presjob)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |-------------------------|
##
##
## Total Observations in Table:  1119
##
##
##              | yvar
##         xvar |     Approve |     Neutral | Not Approve |   Row Total |
## -------------|-------------|-------------|-------------|-------------|
##          Low |          40 |          45 |         284 |         369 |
##              |       75.82 |       13.41 |      127.48 |             |
##              |        0.11 |        0.12 |        0.77 |        0.33 |
## -------------|-------------|-------------|-------------|-------------|
##       Medium |         206 |         121 |         131 |         458 |
##              |        3.86 |        6.64 |       14.55 |             |
##              |        0.45 |        0.26 |        0.29 |        0.41 |
## -------------|-------------|-------------|-------------|-------------|
```

```
##         High |         193 |          68 |          31 |         292 |
##              |       53.72 |        0.79 |       62.64 |             |
##              |        0.66 |        0.23 |        0.11 |        0.26 |
## -------------|-------------|-------------|-------------|-------------|
## Column Total |         439 |         234 |         446 |        1119 |
## -------------|-------------|-------------|-------------|-------------|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  358.9038     d.f. =  4     p =  2.096017e-76
##
##
##
```

```
cross_tab(voters2$female, voters2$presjob)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |-------------------------|
##
##
## Total Observations in Table:  1119
##
##
##              | yvar
##         xvar |     Approve |     Neutral | Not Approve |   Row Total |
## -------------|-------------|-------------|-------------|-------------|
##       Female |         240 |         130 |         223 |         593 |
##              |        0.23 |        0.29 |        0.75 |             |
##              |        0.40 |        0.22 |        0.38 |        0.53 |
## -------------|-------------|-------------|-------------|-------------|
##         Male |         199 |         104 |         223 |         526 |
##              |        0.26 |        0.33 |        0.85 |             |
##              |        0.38 |        0.20 |        0.42 |        0.47 |
## -------------|-------------|-------------|-------------|-------------|
## Column Total |         439 |         234 |         446 |        1119 |
## -------------|-------------|-------------|-------------|-------------|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  2.716166     d.f. =  2     p =  0.2571533
##
```

```
##
##
```

```
cross_tab(voters2$race_white, voters2$presjob)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |-------------------------|
##
##
## Total Observations in Table:  1119
##
##
##              | yvar
##         xvar |     Approve |     Neutral | Not Approve |   Row Total |
## -------------|-------------|-------------|-------------|-------------|
##    Non-White |         179 |          79 |          48 |         306 |
##              |       28.95 |        3.52 |       44.85 |             |
##              |        0.58 |        0.26 |        0.16 |        0.27 |
## -------------|-------------|-------------|-------------|-------------|
##        White |         260 |         155 |         398 |         813 |
##              |       10.90 |        1.33 |       16.88 |             |
##              |        0.32 |        0.19 |        0.49 |        0.73 |
## -------------|-------------|-------------|-------------|-------------|
## Column Total |         439 |         234 |         446 |        1119 |
## -------------|-------------|-------------|-------------|-------------|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  106.4275     d.f. =  2     p =  7.754516e-24
##
##
##
```

```
# Spending Sentiment by Party and Race
cross_tab(voters2$female, voters2$srv_spend)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |-------------------------|
##
```

```
##
## Total Observations in Table:  1119
##
##
##              | yvar
##         xvar |       Low |    Medium |      High | Row Total |
## -------------|-----------|-----------|-----------|-----------|
##       Female |       177 |       265 |       151 |       593 |
##              |      1.76 |      2.05 |      0.09 |           |
##              |      0.30 |      0.45 |      0.25 |      0.53 |
## -------------|-----------|-----------|-----------|-----------|
##         Male |       192 |       193 |       141 |       526 |
##              |      1.98 |      2.31 |      0.10 |           |
##              |      0.37 |      0.37 |      0.27 |      0.47 |
## -------------|-----------|-----------|-----------|-----------|
## Column Total |       369 |       458 |       292 |      1119 |
## -------------|-----------|-----------|-----------|-----------|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  8.289098     d.f. =  2     p =  0.01585058
##
##
##
```

```
cross_tab(voters2$race_white, voters2$srv_spend)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |-------------------------|
##
##
## Total Observations in Table:  1119
##
##
##              | yvar
##         xvar |       Low |    Medium |      High | Row Total |
## -------------|-----------|-----------|-----------|-----------|
##    Non-White |        59 |       149 |        98 |       306 |
##              |     17.40 |      4.51 |      4.13 |           |
##              |      0.19 |      0.49 |      0.32 |      0.27 |
## -------------|-----------|-----------|-----------|-----------|
##        White |       310 |       309 |       194 |       813 |
##              |      6.55 |      1.70 |      1.55 |           |
##              |      0.38 |      0.38 |      0.24 |      0.73 |
## -------------|-----------|-----------|-----------|-----------|
```

```
## Column Total |       369 |       458 |       292 |      1119 |
## -------------|-----------|-----------|-----------|-----------|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  35.83435     d.f. =  2     p =  1.654511e-08
##
##
##
```

Multinomial Logistic Regression Model

- Estimate a multinomial logistic regression with only `age`, `female`, and `race_white` as explanatory variables. Call the regression `mod.nomial1`
- Discussion the estimation results. For instance, is a male more or less likely to be a Democrat (relative to being a Republican)? Answer questions like this using your regression results.

```
# mod.nominal1 <- multinom(FORMULA, data = voters2)
# summary(YOUR ESTIMATED MODEL)
```

Statistical Inference

- As starter, test the existence of the age effect in the logit of independent vs democrat equation. (Hint: For simplicity, use Wald test.)
- Test the existence of effects of an explanatory variable on all response categories.

```
# YOUR CODE TO BE HERE
```

Model Interpretation

- Interpret the estimated coefficients of the model in terms of estimated odds

To interpret the coefficients, we first exponentiate the estimated coefficients

```
# YOUR CODE TO BE HERE
```

Calculation of Estimated Probabilities

- Estimated probabilities for each of the observations in the sample (it's also called "Fitted Value")
- Discuss the estimated probabilities

In practice, however, one could obtaine these estimated probability by simply call the *predict()* function with the correct parameter and a dataset from which the estimated probabilities will be calculated.

```
# YOUR CODE TO BE HERE
```