T-Test Assignment

Zuleyma Castillo

T-test

```
library(ggplot2)
```

```
file.path <- "/cloud/project/Data/Dataset.csv"
Dataset.csv <-read.csv(file.path)
head(Dataset.csv)
```

```
##   S.NoSort.column Case.Identification.Number CONTROL..COMPLETES.1.PHONE.2....
## 1              NA          CASEIDSort column                   MODESort column
## 2               1                           1                         (2) Mail
## 3               2                           2                          (9) Web
## 4               3                           3                          (9) Web
## 5               4                           4                          (9) Web
## 6               5                           5                        (1) Phone
##   CENSUS.STATE.FIPS.CODES.LABE.... Added..Census.Region Added..Census.Division
## 1                STFIPSSort column    REGIONSort column    DIVISIONSort column
## 2                    (51) VIRGINIA            (3) South     (5) South Atlantic
## 3                   (6) CALIFORNIA             (4) West            (9) Pacific
## 4                  (28) MISSISSIPPI            (3) South (6) East South Central
## 5                     (36) NEW YORK        (1) Northeast       (2) Mid-Atlantic
## 6                      (18) INDIANA          (2) Midwest (3) East North Central
##   OFFERED.INTAKE.ASS.MNT.REFER.... Offered.Detox.3.31.03 Offered.SA.Tx.3.31.03
## 1              OTHNONTXSort column      DETOXSort column    TREATMTSort column
## 2                          (1) Yes                (0) No               (1) Yes
## 3                          (1) Yes                (0) No               (1) Yes
## 4                          (1) Yes               (1) Yes               (1) Yes
## 5                           (0) No                (0) No               (1) Yes
## 6                          (1) Yes                (0) No               (1) Yes
##   Setting..Halfway.House OFFERED.OTHER.SA.SERVICES.SU....
## 1        LOC5Sort column                 ADMINSort column
## 2                (1) Yes                         (1) Yes
## 3                (1) Yes                         (1) Yes
## 4                 (0) No                          (0) No
## 5                 (0) No                          (0) No
## 6                 (0) No                          (0) No
##               PRIMARY.FOCUS.SA.TX.MH.MIX.G....
## 1                           FOCUSSort column
## 2       (1) Substance abuse treatment services
## 3       (1) Substance abuse treatment services
## 4 (3) Mix of mental health and substance abuse
## 5       (1) Substance abuse treatment services
## 6 (3) Mix of mental health and substance abuse
##                                   Ownership Federal.Government.Agency
```

```
## 1                        OWNERSHPSort column         FEDOWNSort column
## 2        (2) Private non-profit organization                       <NA>
## 3        (2) Private non-profit organization                       <NA>
## 4        (1) Private for-profit organization                       <NA>
## 5 (4) Local, county, or community government                       <NA>
## 6        (2) Private non-profit organization                       <NA>
##      Solo.practice AFFILIATED.W..RELIGIOUS.ORGA....
## 1 LOC15Sort column                 RELIGSort column
## 2           (0) No                           (0) No
## 3           (0) No                           (0) No
## 4           (0) No                           (0) No
## 5             <NA>                             <NA>
## 6           (0) No                           (0) No
##   LOCATED.IN.OPERATED.BY.HOSPI....            Hospital.Type Hotline...yes.no
## 1              HOSPITALSort column          LOCSSort column HOTYNSort column
## 2                          (0) No                      <NA>           (0) No
## 3                          (0) No                      <NA>           (0) No
## 4                         (1) Yes (2) Psychiatric hospital          (1) Yes
## 5                          (0) No                      <NA>           (0) No
## 6                          (0) No                      <NA>           (0) No
##   Assessment.comprehensive.SA Assessment.mental.health
## 1            SRVC1Sort column         SRVC2Sort column
## 2                       (0) No                   (0) No
## 3                      (1) Yes                     <NA>
## 4                      (1) Yes                  (1) Yes
## 5                      (1) Yes                     <NA>
## 6                      (1) Yes                  (1) Yes
##   Therapy.family.counseling    Therapy.group
## 1          SRVC4Sort column SRVC5Sort column
## 2                     (0) No          (1) Yes
## 3                    (1) Yes          (1) Yes
## 4                    (1) Yes          (1) Yes
## 5                       <NA>          (1) Yes
## 6                    (1) Yes          (1) Yes
```

```r
names(Dataset.csv)
```

```
##  [1] "S.NoSort.column"                  "Case.Identification.Number"
##  [3] "CONTROL..COMPLETES.1.PHONE.2...." "CENSUS.STATE.FIPS.CODES.LABE...."
##  [5] "Added..Census.Region"             "Added..Census.Division"
##  [7] "OFFERED.INTAKE.ASS.MNT.REFER...." "Offered.Detox.3.31.03"
##  [9] "Offered.SA.Tx.3.31.03"            "Setting..Halfway.House"
## [11] "OFFERED.OTHER.SA.SERVICES.SU...." "PRIMARY.FOCUS.SA.TX.MH.MIX.G...."
## [13] "Ownership"                        "Federal.Government.Agency"
## [15] "Solo.practice"                    "AFFILIATED.W..RELIGIOUS.ORGA...."
## [17] "LOCATED.IN.OPERATED.BY.HOSPI...." "Hospital.Type"
## [19] "Hotline...yes.no"                 "Assessment.comprehensive.SA"
## [21] "Assessment.mental.health"         "Therapy.family.counseling"
## [23] "Therapy.group"
```

```r
# Clean FIPS state codes column
Dataset.csv$CENSUS_STATE <- as.numeric(gsub("[^0-9]", "", Dataset.csv$`CENSUS.STATE.FIPS.CODES.LABE....
summary(Dataset.csv$CENSUS_STATE)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    6.00   12.00   29.00   27.56   36.00   72.00       1
```

```r
valid_states <- na.omit(Dataset.csv$CENSUS_STATE)
summary(valid_states)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    6.00   12.00   29.00   27.56   36.00   72.00
```

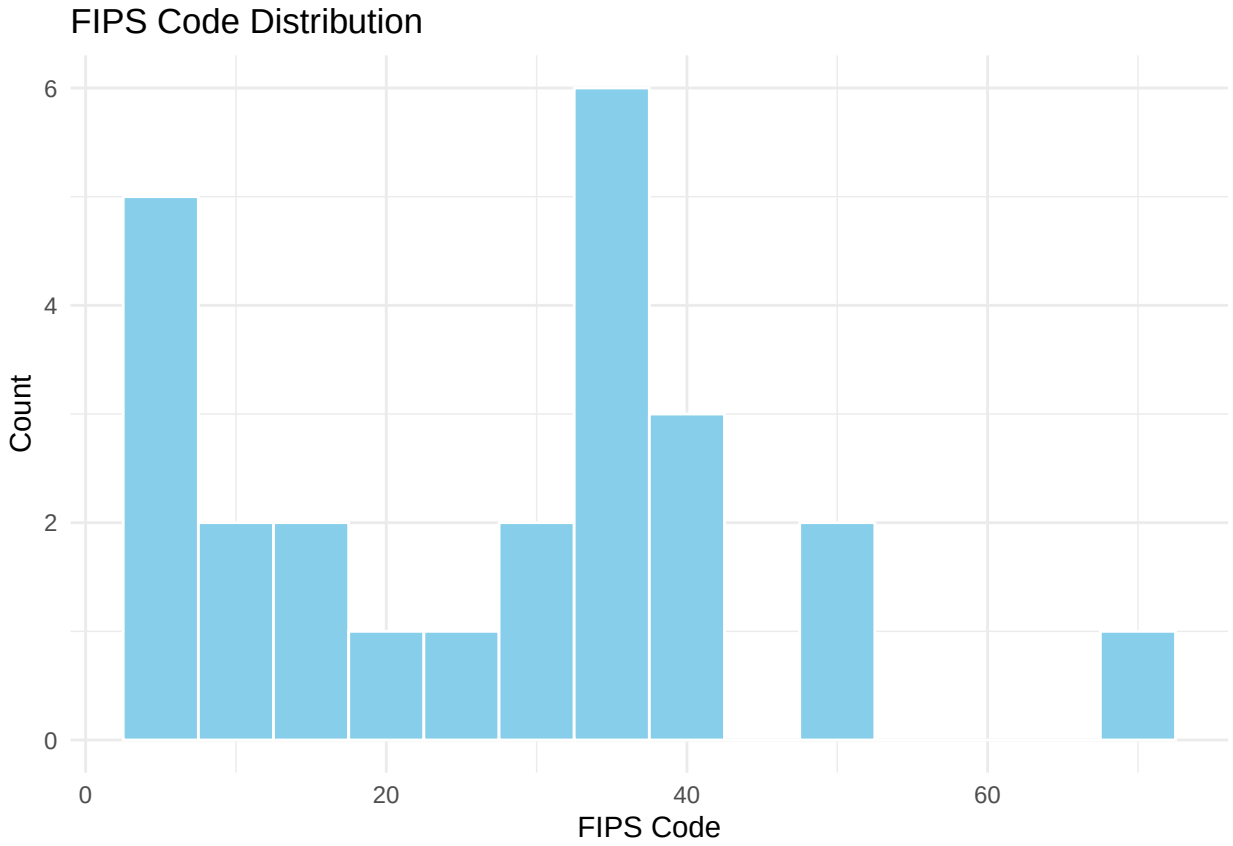
```r
t.test(valid_states, mu = 30, alternative = "two.sided")
```

```
##
##  One Sample t-test
##
## data:  valid_states
## t = -0.70029, df = 24, p-value = 0.4905
## alternative hypothesis: true mean is not equal to 30
## 95 percent confidence interval:
##  20.36878 34.75122
## sample estimates:
## mean of x
##     27.56
```

```r
library(ggplot2)

# Create a data frame for ggplot
df <- data.frame(fips = valid_states)

# Simple ggplot histogram
ggplot(df, aes(x = fips)) +
  geom_histogram(binwidth = 5, fill = "skyblue", color = "white") +
  labs(title = "FIPS Code Distribution",
       x = "FIPS Code",
       y = "Count") +
  theme_minimal()
```



A one-sample t-test was conducted to determine whether the mean of state FIPS codes in the dataset differed significantly from 30. There was some manipulation on variables to convert categorical to numericals in order to complete the assigment

The sample mean was 27.56. The results showed no statistically significant difference from the hypothesized population mean of 30, $t(24) = -0.70$, $p = 0.491$. The 95% confidence interval for the mean ranged from 20.37 to 34.75, which includes the value 30. Therefore, we fail to reject the null hypothesis and conclude that the average FIPS code in this sample does not significantly differ from 30.