Lab 12

Alexander Hernandez

10/6/2022

```
library(PASWR)
```

```
## Loading required package: lattice
```

```
library(MASS)
library(readxl)
```

1) Internet Traffic Data

a) Identify variables and dims

```
Q1 = read.table("http://jse.amstat.org/datasets/packetdata.dat.txt",
                header=TRUE)
```

b) Print frist five observations

```
head(Q1, 5)
```

```
##   timestamp source destination sourceport destport databytes
## 1  0.416754      1           2       1223     2046         0
## 2  0.418705      2           3       1985       20         0
## 3  0.420657      4           5        119     3849         5
## 4  0.426512      3           2         20     1985       512
## 5  0.427488      3           2         20     1985       512
```

c) Is there any missing value?

```
# For count: sum(is.na(Q1))
any(is.na(Q1))
```

```
## [1] FALSE
```

d) Construct a 90% conf int for the average timestamp

```
t.test(Q1$time, conf.leve=0.9)$conf.int
```

```
## [1] 50.00747 50.42834
## attr(,"conf.level")
## [1] 0.9
```

2) Maternal Smoking Impact on Infant Health

a) Import the data

```
smoking = read.table("http://www.stat.berkeley.edu/~statlabs/data/babiesI.data",
                     header=TRUE)
```

b) How many observations have smoking status unknown

```
nrow(smoking[smoking$smoke == "9",])
```

```
## [1] 10
```

```
# Option 2 :table(smoking$smoke)
```

c) Clean dataset by removing subjects with unknown smoking status

```
new_smoking = subset(smoking, smoking$smoke!="9")
dim(smoking)
```

```
## [1] 1236    2
```

```
dim(new_smoking)
```

```
## [1] 1226    2
```

d) Is there evidence that the newborn baby will have significantly low weight for a smoker mom than for a non-smoker mom?

```
# Null: u1 - u2 = 0
# Alt:  u1 - u2 > 0
t.test(new_smoking$bwt ~ new_smoking$smoke, alt = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  new_smoking$bwt by new_smoking$smoke
## t = 8.5813, df = 1003.2, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group 0 and group 1 is greater than 0
## 95 percent confidence interval:
##  7.222928      Inf
## sample estimates:
## mean in group 0 mean in group 1
##        123.0472        114.1095
```

```
# With p-value of 2.2e-16,
# we have enough evidence to reject the null hypothesis and
# can claim that newborn babies have lower weights with smoking mothers.
```

3) Bottle Water

```
# Null: u1 - u2 = 0
# Alt:  u1 - u2 < 0
t.test(Water$Sodium ~ Water$Source, alt="less")
```

```
##
##  Welch Two Sample t-test
##
## data:  Water$Sodium by Water$Source
## t = -1.8589, df = 22.069, p-value = 0.03822
## alternative hypothesis: true difference in means between group X and group Y is less than 0
## 95 percent confidence interval:
##        -Inf -0.3665724
## sample estimates:
## mean in group X mean in group Y
##            76.4            81.2
```

```
# With a p-value of 0.03822,
# we have enough evidence to reject the null hypothesis and
# can claim that there is less sodium in source X than Y.
```

4) Foot Measurements of Fourth Graders

```
# Null: u(b) - u(g) = 0
# Alt:  u(b) - u(g) > 0
feet = read.table("http://ww2.amstat.org/publications/jse/datasets/kidsfeet.dat.txt",
                  header=FALSE)
t.test(feet$V2 ~ feet$V5)
```

```
##
##  Welch Two Sample t-test
##
## data:  feet$V2 by feet$V5
## t = -1.1778, df = 34.309, p-value = 0.247
## alternative hypothesis: true difference in means between group B and group G is not equal to 0
## 95 percent confidence interval:
##  -0.3943871  0.1049134
## sample estimates:
## mean in group B mean in group G
##        87.75000        87.89474
```

```
# With a p-value of 0.247,
# we do not have enough evidence to reject the null hypothesis
# and cannot claim there is a difference in feet length of
# fourth graders by gender.
```

5) Capital Punishment under 18 Poll

```
# Null: u(c) - u(s)  = 0
# Alt:  u(c) - u(s) != 0
prop.test(c(180,238), c(580,600), correct=F)
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  c(180, 238) out of c(580, 600)
## X-squared = 9.6066, df = 1, p-value = 0.001939
```

```
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.14063404 -0.03200964
## sample estimates:
##    prop 1    prop 2
## 0.3103448 0.3966667
```

```
# With a p-value of 0.0023 (or 0.0019 without correction),
# we have enough evidence to reject the null hypothesis
# and claim that the proportion between Catholics and seculars is different
```

6) Female Hurricanes vs Male Hurricanes

```
# Null: u(f) - u(m) = 0
# Alt:  u(f) - u(m) > 0
hurricanes1 = read_xlsx("C:\\repos\\STAT 50001\\Lab 12\\Hurricane.xlsx",
                        range="A1:D47")
hurricanes2 = read_xlsx("C:\\repos\\STAT 50001\\Lab 12\\Hurricane.xlsx",
                         range="E1:H47")
hurricanes = rbind(hurricanes1, hurricanes2)

t.test(hurricanes$Death[hurricanes$Gender=="Female"],
       hurricanes$Death[hurricanes$Gender=="Male"],
       alt="greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  hurricanes$Death[hurricanes$Gender == "Female"] and hurricanes$Death[hurricanes$Gender == "Ma
## t = 1.9022, df = 86.161, p-value = 0.03024
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  1.622587      Inf
## sample estimates:
## mean of x mean of y
##  24.71429  11.82759
```

```
# With a p-value of 0.03024,
# we have enough evidence to reject the null hypothesis
# and claim can claim that female hurricanes cause more deaths than male ones.
```

7) birthwt data and smoking

```
# Null: u(s) - u(ns) = 0
# Alt:  u(s) - u(ns) > 1

attach(birthwt)
table(low)
```

```
## low
##   0   1
## 130  59
```

```
xtabs(~low+smoke)
```

```
##    smoke
## low  0  1
##   0 86 44
##   1 29 30
```

```
prop.test(c(29,30), c(115,74))
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(29, 30) out of c(115, 74)
## X-squared = 4.2359, df = 1, p-value = 0.03958
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.301495793 -0.004967192
## sample estimates:
##    prop 1    prop 2
## 0.2521739 0.4054054
```

```
# With a p-value of 0.0396,
# we have enough evidence to reject the null hypothesis
# and can claim that there is a higher fraction of low-birth weights
# in smoking mothers versus non-smoking mothers.
```