Homework2_Hwang

Charles Hwang

1/29/2022

Charles Hwang

Dr. Perry

STAT 451-001

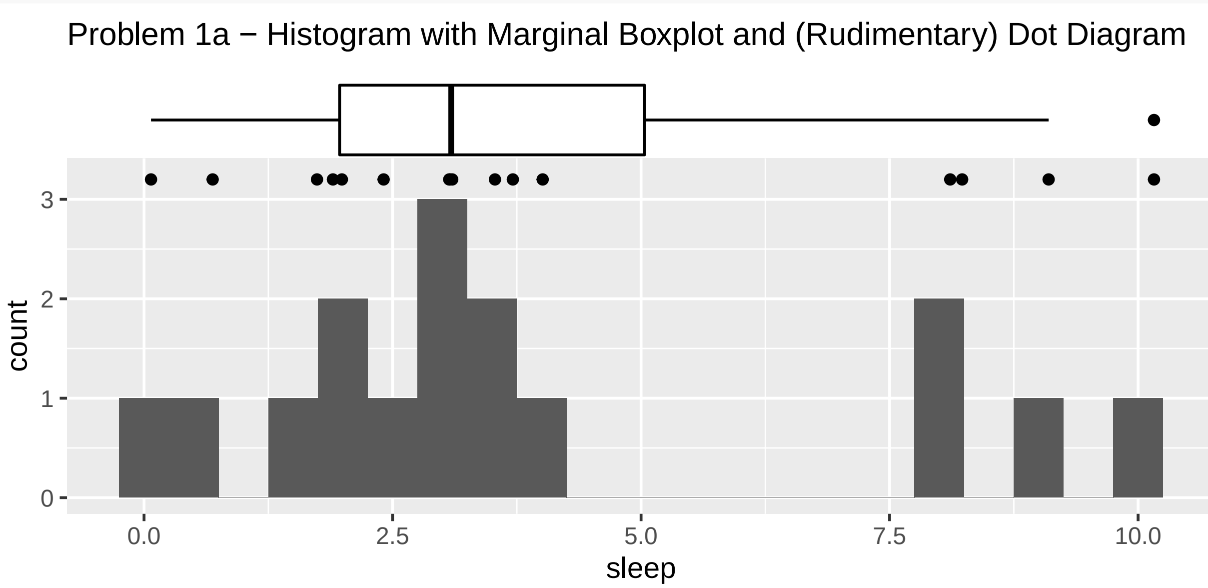
29 January 2022

Problem 1

```
rm(list=ls())
library(ggplot2)
library(ggExtra)
library(pwr)
library(binom)
sleep<-sort(c(1.90,3.08,9.10,3.53,1.99,3.10,10.16,0.69,1.74,2.41,4.01,3.71,8.11,8.23,0.07,3.07))
analysis<-data.frame(mean(sleep),median(sleep),sd(sleep),max(sleep)-min(sleep)) # Problem 1a
names(analysis)<-c("Mean","Median","Std. Dev.","Range")
analysis
```

```
##      Mean Median Std. Dev. Range
## 1 4.05625   3.09  3.098421 10.09
```

```
ggMarginal(ggplot(data.frame(sleep),aes(x=sleep))+geom_point(y=3.2)+geom_histogram(binwidth=0.5)+coord_
```



A nonparametric test may be better for this data because they do not appear to be normally distributed.

```r
pbinom(13,length(sleep),0.5)-pbinom(4-1,length(sleep),0.5)                    # Problem 1b
```

```
## [1] 0.9872742
```

```r
sleep[c(qbinom(0.005,length(sleep),0.5),qbinom(0.995,length(sleep),0.5))]     # Problem 1c
```

```
## [1] 1.74 8.11
```

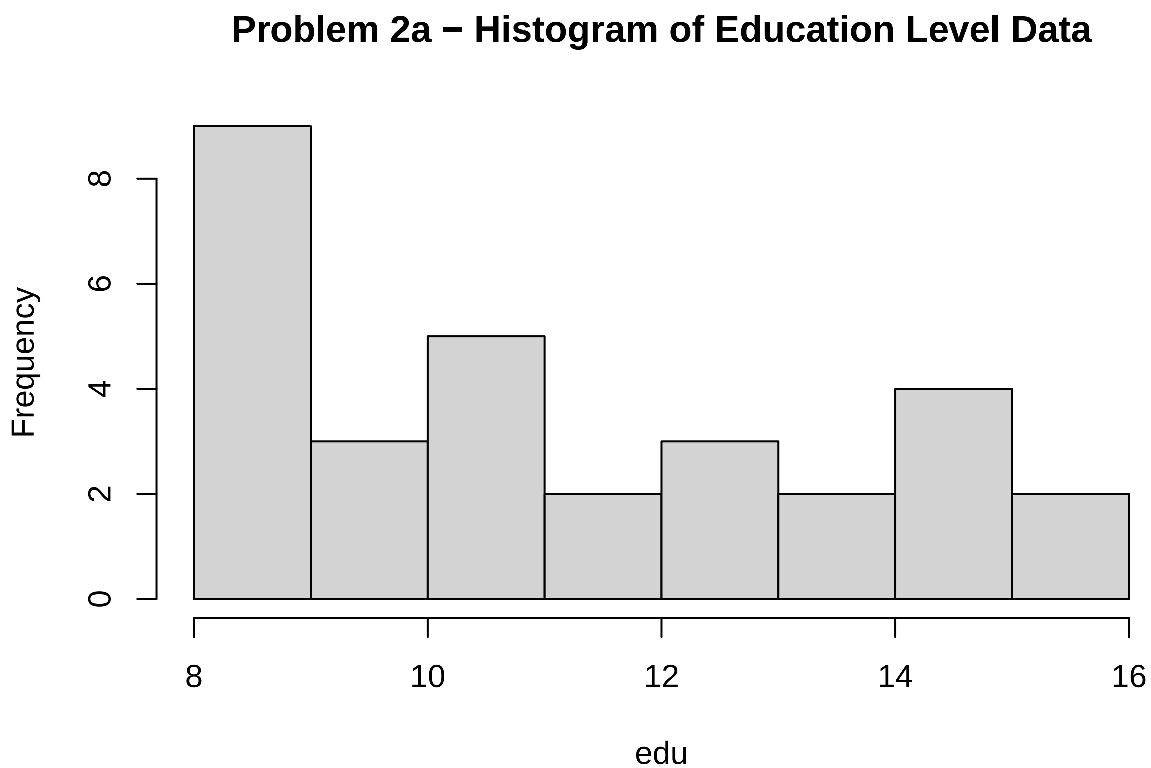
```r
# We can see these are the 3rd and 13th values in the dataset sorted in ascending order.
pbinom(13,length(sleep),0.5)-pbinom(3-1,length(sleep),0.5)                    # Problem 1d
```

```
## [1] 0.9958191
```

We are approximately 99.5819092 percent confident that the true *median* percentage of sleep time spent in stage 0 for healthy males aged 50-60 is between 1.74 and 8.11 percent.

Problem 2

```r
edu<-read.table("/Users/newuser/Desktop/Notes/Graduate/STAT 451 - Nonparametric Statistical Methods/Edu
edu<-sort(as.numeric(edu$Educ_Level))
hist(edu,main="Problem 2a - Histogram of Education Level Data") # Problem 2a
```




```r
# The data appear to be right-skew and not normally distributed.
# H0: mu = 12.4 years                                                 # Problem 2b
# HA: mu < 12.4 years
t.test(edu,mu=12.4,alternative="less")
```

```
##
##  One Sample t-test
##
## data:  edu
## t = -1.8889, df = 29, p-value = 0.03447
```

```
## alternative hypothesis: true mean is less than 12.4
## 95 percent confidence interval:
##      -Inf 12.30956
## sample estimates:
## mean of x
##      11.5
```

```
# We reject H0 at the alpha = 0.05 level. There is sufficient evidence (p = 0.03447)
# that the mean education level is less than 12.4 years.
pwr.t.test(d=(mean(edu)-12.4)/(sd(edu)/sqrt(length(edu))),n=length(edu),type="one.sample",alternative="l
```

```
##
##      One-sample t test power calculation
##
##               n = 30
##               d = -1.888942
##       sig.level = 0.05
##           power = 1
##     alternative = less
```

```
median(edu) # We can see the power of this test is approximately 1 - beta = 1.00.
```

```
## [1] 11
```

```
# H0: m = 11                                                    # Problem 2c
# HA: m < 11
table(sign(edu-median(edu)))
```

```
##
## -1  0  1
## 12  5 13
```

```
# We split the ties among positive and negative as evenly as possible, leaving 12 + 2 = 14
binom.test(14,30,alternative="less") # negatives, 5 - 2*2 = 1 tie, and 13 + 2 = 15 positives.
```

```
##
##  Exact binomial test
##
## data:  14 and 30
## number of successes = 14, number of trials = 30, p-value = 0.4278
## alternative hypothesis: true probability of success is less than 0.5
## 95 percent confidence interval:
##  0.0000000 0.6300524
## sample estimates:
## probability of success
##              0.4666667
```

```
# We fail to reject H0 at the alpha = 0.05 level. There is insufficient
# evidence (p = 0.4278) that the median is less than 11.
# H0: m = 11                                                    # Problem 2d
# HA: m < 11
pnorm((14-length(edu)*0.5)/sqrt(length(edu)*0.5*(1-0.5)))
```

```
## [1] 0.3575003
```

```
# We fail to reject H0 at the alpha = 0.05 level. There is insufficient
# evidence (p = 0.3575003) that the median is less than 11.
pwr.norm.test(d=(14-length(edu)*0.5)/sqrt(length(edu)*0.5*(1-0.5)),n=length(edu),alternative="less")
```

```
##
##      Mean power calculation for normal distribution with known variance
##
##               d = -0.3651484
##               n = 30
##       sig.level = 0.05
##           power = 0.63876
##     alternative = less
```

We can see the power for the one-proportion z-test is approximately $1 - \beta = 0.63876$.

Problem 2e

I believe the one-sample t-test is the most appropriate for this data. Although we assume the population distribution is symmetric, looking at the histogram in Problem 2a, we cannot assume the same of this sample data. Since $n = 30$, the Central Limit Theorem is generally applicable to this data.

Problem 3

```
prop.test(2,30)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  2 out of 30, null probability 0.5
## X-squared = 20.833, df = 1, p-value = 5.01e-06
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.01163184 0.23507287
## sample estimates:
##          p
## 0.06666667
```

```
binom.confint(x=2,n=30,method=c("wilson","ac","asymptotic"))
```

```
##          method x  n       mean        lower     upper
## 1 agresti-coull 2 30 0.06666667  0.008024712 0.2236869
## 2    asymptotic 2 30 0.06666667 -0.022594020 0.1559274
## 3        wilson 2 30 0.06666667  0.018477024 0.2132346
```

We can see the proportion test method has the widest confidence interval at (0.0116, 0.2351), while the asymptotic method has the most conservative interval at (-0.0226, 0.1559). It is interesting to see how the Agresti-Coull and Wilson's adjusted intervals are not centered around $\bar{x} = \frac{2}{30}$ as one might initially think; instead, they are both centered around 0.1158558, which may be due to $\bar{x}$ being so close to 0.