

Lab__1

Frank Woodling

February 4, 2016

Results

1.

We fail to reject the null hypothesis. We cannot conclude that the average smoker started smoking before the age of 18.

2.

It is hard to tell the distribution with the default settings as there are only 4 bins. Even with smaller breaks the one age 30 seems to throw off the distribution. It would be hard to assume this data is normal, however, you it does look normal-ish with the center on 17.

3.

We can reject the null hypothesis in this case. We can conclude than the age people started smoking is less than 18.

4.

We were able to reject on the binomial test, but we failed to reject on the t-test. With such a small sample size the one outlier at age 30 seems to throw off the t-test. In this case the binomial test may be better since it is more robust.

5.

`SIGN.test(smokeage, md=18, alternative="less")` Code is in appendix.

6.

A confidence bound of 17.2 means that 95% of the population is less than 17.2.

7.

The 99% confidence interval goes from 13.8 to 19. THis means that out of an infinite number of confidence intervals 99% of them will contain the true mean.

8.

For the two test using “less” and “greater” as the alternative there is only a confidence bound, and not an interval. In the two.sided test the lower bound of the interval is lower than the confidence bound in the “less” test. The same is true for the upper bound; the upper bound of the interval in the two sided test is greater than the upper bound in the “greater” only test.

We can reject the null hypothesis on the “less” test. We fail to reject the null hypothesis on the “greater” test. We also fail to reject the null on the two-sided test.

9.

The mean and median are both very close to 10. The two are very similar. The histogram looks mostly normal. It may be helpful to adjust bin sizes.

The p-value for the t-test is 7.571e-06. We can reject the null hypothesis that the true mean is greater than 8.

The p-value for the sign test is 0.0004681. We can also reject the null hypothesis in the binomial test. The binomial test has a slightly smaller lower bound, though they are very similar. Both p-values were also very small.

10.

Looking at the histogram the distribution of the data is clearly right skewed. The data is not normally distributed. We can also see that the median and mean are very far from each other (8.2 median, 9.9 mean).

The p-value for the t-test is 0.0703, so we fail to reject the null hypothesis.

The p-value for the sign test is 0.4439, so we fail to reject the null hypothesis.

The two tests give much different looking results compared to the previous data set.

We fail to reject the null hypothesis in both tests. The confidence bounds are very far apart through, which the t-test at 7.8 and the sign-test at 5.0.

Summary

We ran two different tests on the smoking data. One test we were able to reject the null hypothesis and the other we were not. The data set is very small, so a single extreme value can skew the results of a t-test easily. In this case the binomial test is the better option. Since it uses medians it is more robust, and less affected by a single outlier. Using the binomial test we were able to reject the null hypothesis, concluding the average smoker started smoking before the age of 18.

Code

```
# Part 1.
smokeage = c(18,19,30,16,17,15,14,14,17,12,14,13,19,19,17,13,20,12,17,15)
stddev <- sqrt(sum((smokeage-mean(smokeage))^2)/(length(smokeage)-1))
t <- ((mean(smokeage)-18)/stddev)*sqrt(length(smokeage))
t
```

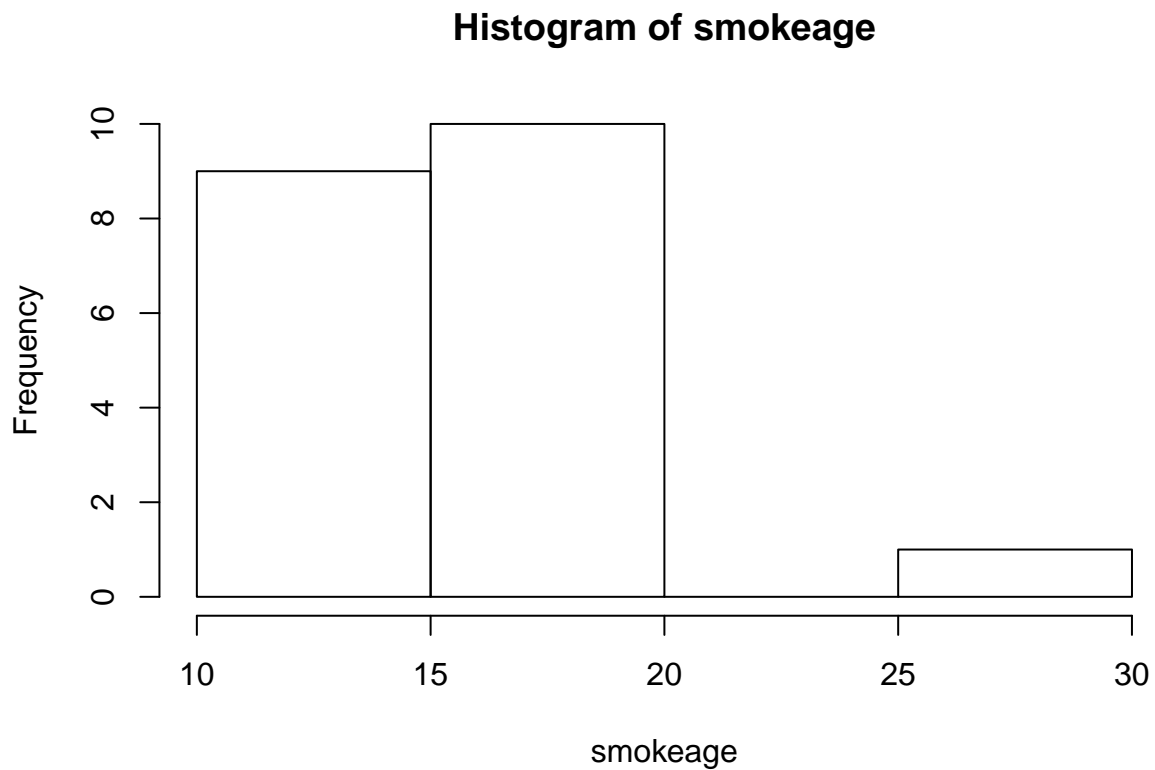
```
## [1] -1.618622
```

```
# [1] -1.618622  
pt(t, df=length(smokeage)-1)
```

```
## [1] 0.06100454
```

```
# [1] 0.06100454  
# We fail to reject the null hypothesis. We cannot conclude that the average  
# smoker started smoking before the age of 18.
```

```
# Part 2.  
hist(smokeage)
```



```
hist(smokeage, breaks=(10:30))
```

```
# It is hard to tell the distribution with the default settings as there are only 4 bins.  
# even with smaller breaks the one age 30 seems to throw off the distribution. It would be  
# hard to assume this data is normal, however, you it does look normal-ish with the center on 17.
```

```
# Part 3.  
b <- length(subset(smokeage, smokeage <= 17))
```

```
z_b <- (b-0.5*length(smokeage))/sqrt(.25*length(smokeage))
1-pnorm(z_b)
```

```
## [1] 0.03681914
```

```
# [1] 0.03681914
```

```
# We can reject the null hypothesis in this case. We can conclude than the age  
# people started smoking is less than 18.
```

```
# Part 4.  
# We were able to reject on the binomial test, but we failed to reject on the t-test.  
# With such a small sample size the one outlier at age 30 seems to throw off the t-test.  
# In this case the binomial test may be better since it is more robust.
```

```
#####
```

```
# Part 5.
```

```
library('BSDA')
```

```
## Warning: package 'BSDA' was built under R version 3.2.3
```

```
## Loading required package: e1071
```

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

```
##
```

```
## Attaching package: 'BSDA'
```

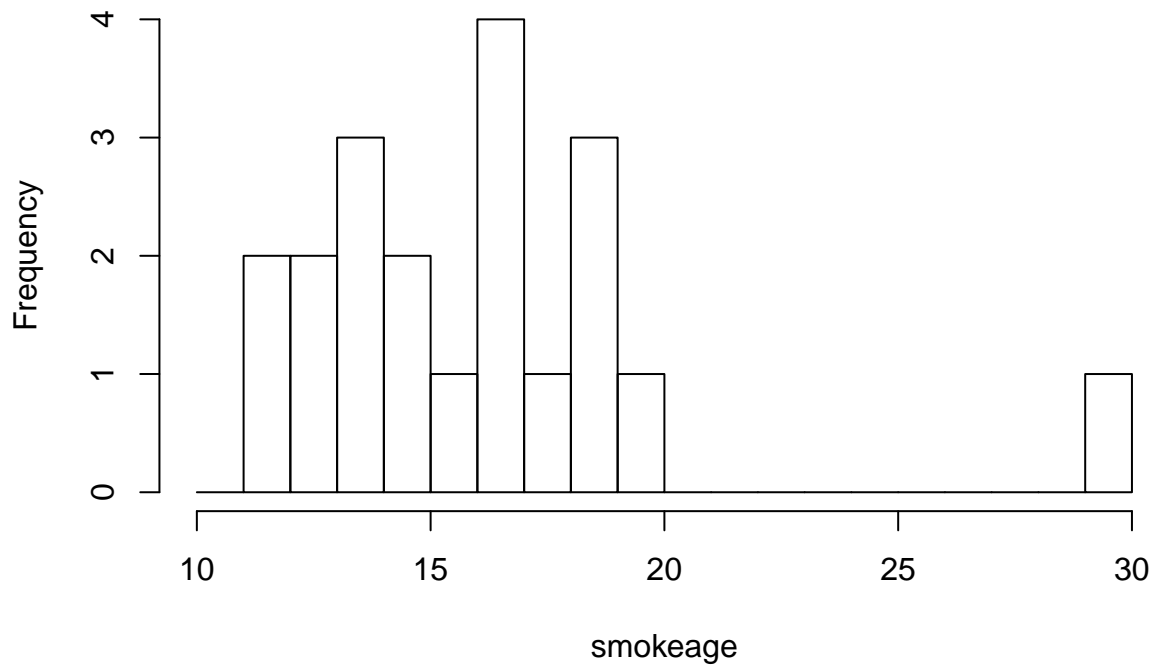
```
##
```

```
## The following object is masked from 'package:datasets':
```

```
##
```

```
## Orange
```

Histogram of smokeage



```
SIGN.test(smokeage, md=18, alternative="less")
```

```
##
## One-sample Sign-Test
##
## data: smokeage
## s = 5, p-value = 0.03178
## alternative hypothesis: true median is less than 18
## 95 percent confidence interval:
## -Inf 17.2072
## sample estimates:
## median of x
## 16.5
```

```
##               Conf.Level L.E.pt U.E.pt
## Lower Achieved CI    0.9423  -Inf 17.0000
## Interpolated CI      0.9500  -Inf 17.2072
## Upper Achieved CI     0.9793  -Inf 18.0000
```

Part 6.

A confidence bound of 17.2 means that 95% of the population is less than 17.2.

Part 7.

```
SIGN.test(smokeage, md=18, alternative="two.sided", conf.level=.99)
```

```
##
## One-sample Sign-Test
##
## data: smokeage
## s = 5, p-value = 0.06357
## alternative hypothesis: true median is not equal to 18
## 99 percent confidence interval:
## 13.80328 19.00000
## sample estimates:
## median of x
## 16.5
```

```
##               Conf.Level L.E.pt U.E.pt
## Lower Achieved CI    0.9882 14.0000    19
## Interpolated CI      0.9900 13.8033    19
## Upper Achieved CI    0.9974 13.0000    19
```

The 99% confidence interval goes from 13.8 to 19. This means that out of an infinite number of confidence intervals 99% of them will contain the true mean.

Part 8.

```
SIGN.test(smokeage, md=18, alternative="less", conf.level=.99)
```

```
##
## One-sample Sign-Test
##
## data: smokeage
## s = 5, p-value = 0.03178
## alternative hypothesis: true median is less than 18
## 99 percent confidence interval:
## -Inf 18.72331
## sample estimates:
## median of x
## 16.5
```

```
##               Conf.Level L.E.pt U.E.pt
## Lower Achieved CI    0.9793 -Inf 18.0000
## Interpolated CI      0.9900 -Inf 18.7233
## Upper Achieved CI    0.9941 -Inf 19.0000
```

```
#               Conf.Level L.E.pt U.E.pt
# Lower Achieved CI    0.9793 -Inf 18.0000
# Interpolated CI      0.9900 -Inf 18.7233
# Upper Achieved CI    0.9941 -Inf 19.0000
```

```
SIGN.test(smokeage, md=18, alternative="greater", conf.level=.99)
```

```
##
## One-sample Sign-Test
##
## data: smokeage
```

```
## s = 5, p-value = 0.9904
## alternative hypothesis: true median is greater than 18
## 99 percent confidence interval:
##    14 Inf
## sample estimates:
## median of x
##        16.5
```

```
##                Conf.Level L.E.pt U.E.pt
## Lower Achieved CI    0.9793    14    Inf
## Interpolated CI      0.9900    14    Inf
## Upper Achieved CI    0.9941    14    Inf
```

```
#                Conf.Level L.E.pt U.E.pt
# Lower Achieved CI    0.9793    14    Inf
# Interpolated CI      0.9900    14    Inf
# Upper Achieved CI    0.9941    14    Inf
```

```
SIGN.test(smokeage, md=18, alternative="two.sided", conf.level=.99)
```

```
##
## One-sample Sign-Test
##
## data:  smokeage
## s = 5, p-value = 0.06357
## alternative hypothesis: true median is not equal to 18
## 99 percent confidence interval:
##  13.80328 19.00000
## sample estimates:
## median of x
##        16.5
```

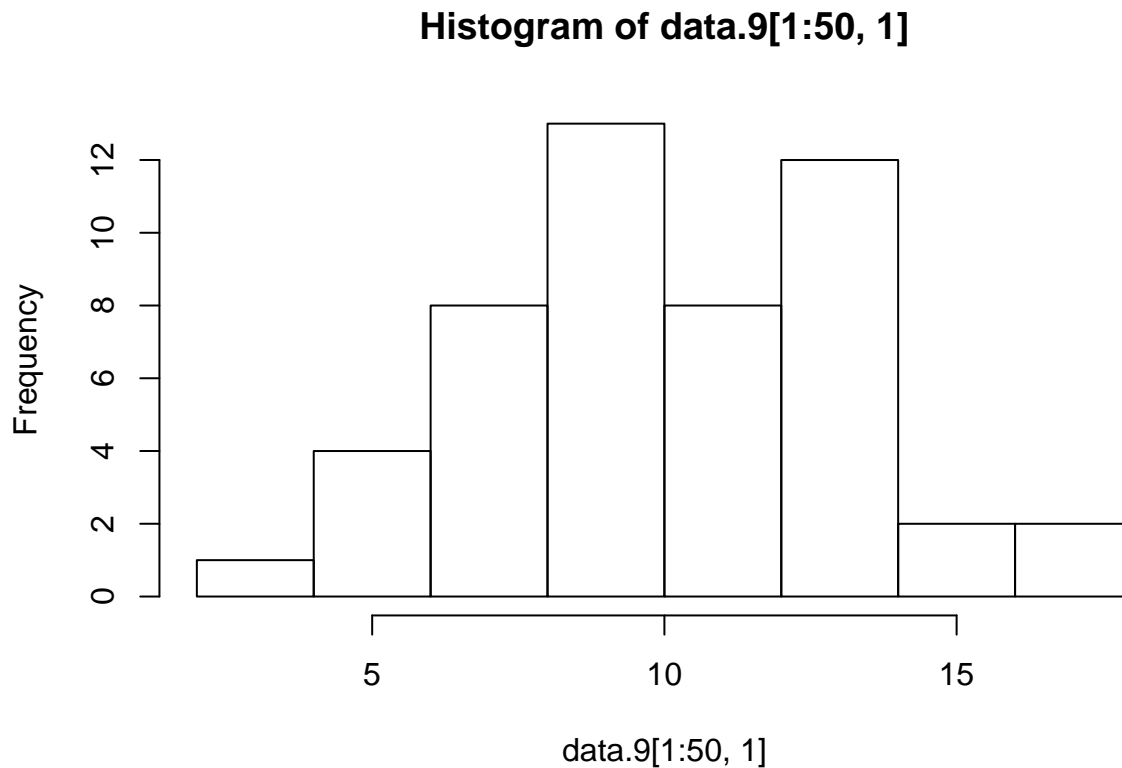
```
##                Conf.Level  L.E.pt U.E.pt
## Lower Achieved CI    0.9882 14.0000    19
## Interpolated CI      0.9900 13.8033    19
## Upper Achieved CI    0.9974 13.0000    19
```

```
#                Conf.Level  L.E.pt U.E.pt
# Lower Achieved CI    0.9882 14.0000    19
# Interpolated CI      0.9900 13.8033    19
# Upper Achieved CI    0.9974 13.0000    19
```

```
# For the two test using "less" and "greater" as the alternative there is only a
# confidence bound, and not an interval. In the two.sided test the lower bound
# of the interval is lower than the confidence bound in the "less" test. The same
# is true for the upper bound; the upper bound of the interval in the two sided test
# is greater than the upper bound in the "greater" only test.
#
# We can reject the null hypothesis on the "less" test. We fail to reject the
# null hypothesis on the "greater" test. We also fail to reject the null on the two-sided
# test.
```

```
#####
# Part 9.

data.9 <- read.table('data.symm.txt', header=T)
hist(data.9[1:50,1])
```



```
summary(data.9[1:50,1])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  3.977  8.064   9.950  10.080  12.510  16.460
```

```
# Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
# 3.977  8.064   9.950  10.080  12.510  16.460
```

```
# The mean and median are both very close to 10. The two are very similar. The histogram
# looks mostly normal. It may be helpful to adjust bin sizes.
```

```
t.test(data.9[1:50,1], mu=8, alternative="greater")
```

```
##
## One Sample t-test
##
## data:  data.9[1:50, 1]
## t = 4.8032, df = 49, p-value = 7.571e-06
```



```
## alternative hypothesis: true mean is greater than 8
## 95 percent confidence interval:
##  9.350882      Inf
## sample estimates:
## mean of x
## 10.07525
```

The p-value is 7.571e-06. We can reject the null hypothesis that the true mean is greater than 8.

```
SIGN.test(data.9[1:50,1], md=8, alternative="greater")
```

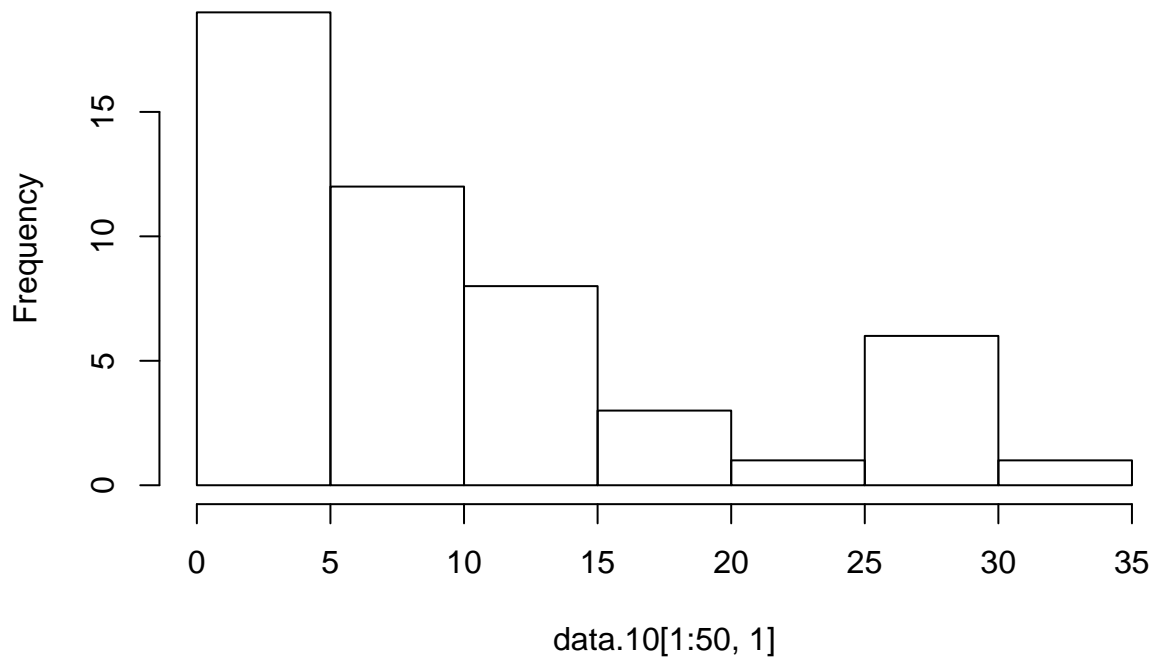
```
##
## One-sample Sign-Test
##
## data: data.9[1:50, 1]
## s = 37, p-value = 0.0004681
## alternative hypothesis: true median is greater than 8
## 95 percent confidence interval:
##  9.278101      Inf
## sample estimates:
## median of x
##  9.950267
```

```
##              Conf.Level L.E.pt U.E.pt
## Lower Achieved CI    0.9405 9.2892   Inf
## Interpolated CI      0.9500 9.2781   Inf
## Upper Achieved CI    0.9675 9.2576   Inf
```

*# The p-value is 0.0004681. We can also reject the null hypothesis in the binomial test.
 # The binomial test has a slightly smaller lower bound, though they are very similar.
 # Both p-values were also very small.*

```
# Part 10.
data.10 <- read.table('data.skew.txt', header=T)
hist(data.10[1:50,1])
```

Histogram of data.10[1:50, 1]



```
summary(data.10[1:50,1])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1847  2.4430   8.2610   9.8660 13.6300 30.8900
```

```
#   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#  0.1847  2.4430   8.2610   9.8660 13.6300 30.8900
```

```
# Looking at the histogram the distribution of the data is clearly right skewed. The data
# is not normally distributed. We can also see that the median and mean are very far from
# each other (8.2 median, 9.9 mean).
```

```
t.test(data.10[1:50,1], mu=8, alternative="greater")
```

```
##
## One Sample t-test
##
## data:  data.10[1:50, 1]
## t = 1.4978, df = 49, p-value = 0.0703
## alternative hypothesis: true mean is greater than 8
## 95 percent confidence interval:
##  7.777324      Inf
## sample estimates:
## mean of x
##  9.865746
```

```
# The p-value is 0.0703, so we fail to reject the null hypothesis.
```

```
SIGN.test(data.10[1:50,1], md=8, alternative="greater")
```

```
##  
## One-sample Sign-Test  
##  
## data: data.10[1:50, 1]  
## s = 26, p-value = 0.4439  
## alternative hypothesis: true median is greater than 8  
## 95 percent confidence interval:  
## 5.022828 Inf  
## sample estimates:  
## median of x  
## 8.261484
```

```
##  
## Conf.Level L.E.pt U.E.pt  
## Lower Achieved CI 0.9405 5.0364 Inf  
## Interpolated CI 0.9500 5.0228 Inf  
## Upper Achieved CI 0.9675 4.9976 Inf
```

```
# The p-value is 0.4439, so we fail to reject the null hypothesis.
```

```
# The two tests give much different looking results compared to the previous data set.  
# We fail to reject the null hypothesis in both tests. The confidence bounds are very  
# far apart through, which the t-test at 7.8 and the sign-test at 5.0.
```

```
# Summary
```

```
# We ran two different tests on the smoking data. One test we were able to reject the null hypothesis  
# and the other we were not. The data set is very small, so a single extreme value can skew the result  
# of a t-test easily. In this case the binomial test is the better option. Since it uses medians  
# it is more robust, and less affected by a single outlier. Using the binomial test we were able to  
# reject the null hypothesis, concluding the average smoker started smoking before the age of 18.
```