Homework 2
t Tests

Steve Bronder
Statistical Inference

October 28, 2014

Exercise 1. Find a dataset and perform a t-test to evaluate whether the mean number of days below 32 degrees for the six weeks after Punxsutawney Phil sees his shadow is greater than or equal to the mean number of days below 32 degrees for the six weeks after Phil does not see his shadow.

Answer

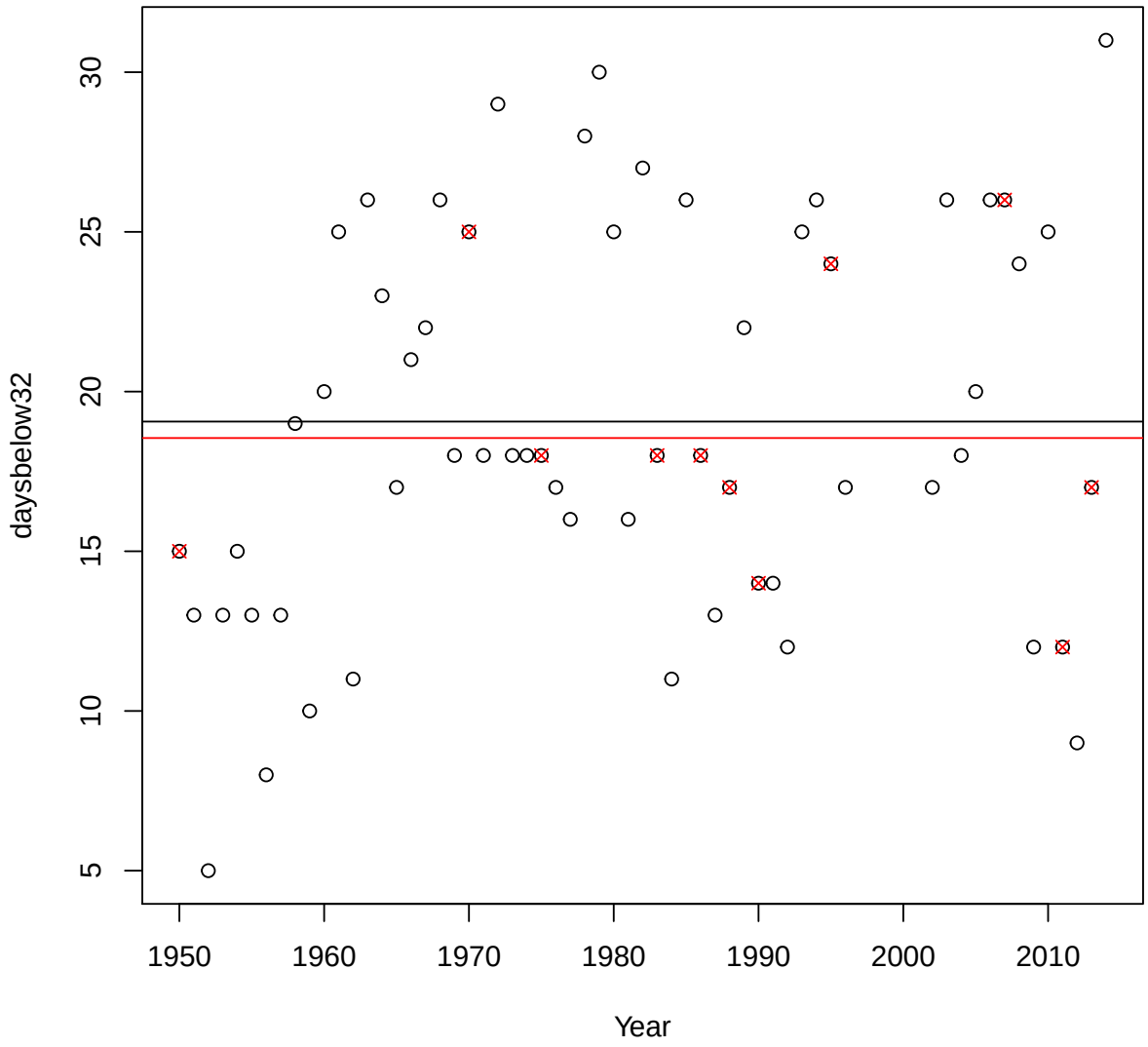
To evaluate whether Punxsutawney Phil can accurately predict six more weeks of winter we have to define a criteria for evaluation. Our criteria will be whether or not Phil accurately predicted a greater than average number of days with temperature below 32 degrees for the months of February and March of each year. If Phil predicts a long winter for a year we will give him 1 point out of a possible 65 points. Phil is limited to 65 points because our data ranges from 1950 to 2014 ($T = 65$). We gather data from the National Climate Data Center[1] for the county Punxsutawney Phil is located in, Jefferson County, PA. To find whether Phil saw his shadow we gather data from stormfax[2] on whether or not he saw his shadow for our target years.

Our data set is comprised of three variables, The year, the number of days it is below 32 degrees in the months of February and March, and whether or not Phil saw his shadow for a given year. Lets have a look at our data. We remove the years in which a moving average was used to gain the number of days for unreported periods. We then plot the data using

```
# Read in csv
Ghog<-read.csv("./ghog_clean.csv",header=TRUE)

# Remove moving average years
Ghog2 <- Ghog[c(1:47,53:65),]
```

[1] http://www.ncdc.noaa.gov/cdo-web/datatools/findstation
[2] http://www.stormfax.com/ghogday.htm

```r
attach(Ghog2)

plot(Year,daysbelow32 )

# No shadow line and shadow line in graph
abline(mean(daysbelow32[ghog=="NOShadow"]),0,col="red")
abline(mean(daysbelow32[ghog=="Shadow"]),0)

# make color points
points(Year[ghog=="NOShadow"],daysbelow32[ghog=="NOShadow"],col="red",pch=4)
```



```r
mean.b32.ns <- mean(daysbelow32[ghog=="NOShadow"])
mean.b32.s <- mean(daysbelow32[ghog=="Shadow"])
sd.b32.ns <- sd(daysbelow32[ghog=="NOShadow"])
sd.b32.s <- sd(daysbelow32[ghog=="Shadow"])
ln.b32.ns <- length(daysbelow32[ghog=="NOShadow"])
ln.b32.s <- length(daysbelow32[ghog=="Shadow"])
```

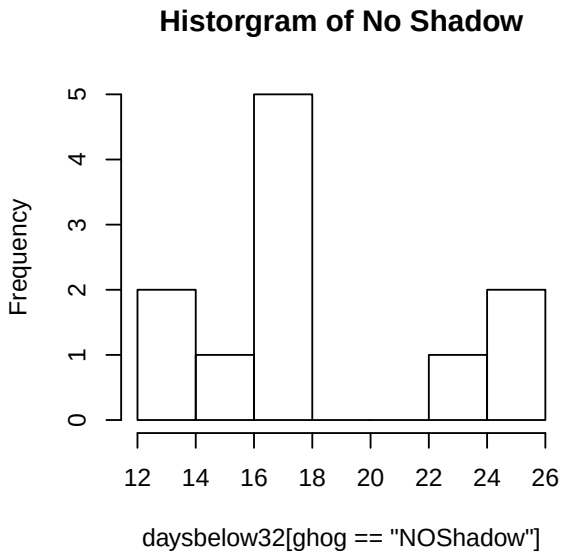
At this point we run a two sample t-test by collecting the mean, standard deviation, and length of the days the temperature was below 32 degrees catagorized by whether or not the almighty groundhog saw his shadow.

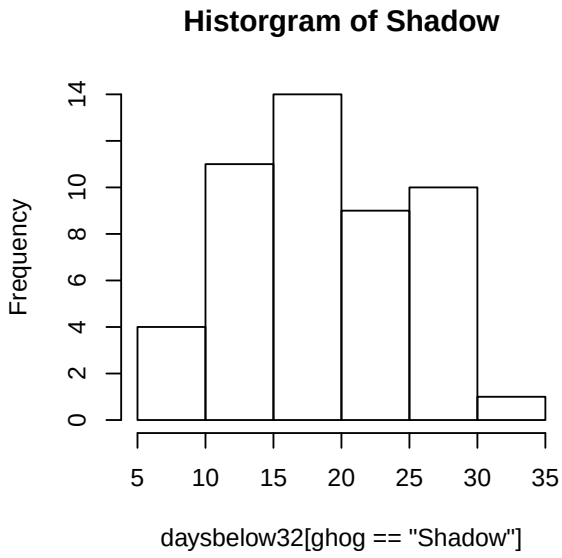
```r
ttest.s <- (mean.b32.ns-mean.b32.s)/sqrt((sd.b32.ns^2/ln.b32.ns) +
                                         (sd.b32.s^2/ln.b32.s))

ttest.s

## [1] -0.3112187

#t test with alternative greater
t.test(daysbelow32[ghog=="NOShadow"],daysbelow32[ghog=="Shadow"],
       alternative=c("greater"))

##
##  Welch Two Sample t-test
##
## data:  daysbelow32[ghog == "NOShadow"] and daysbelow32[ghog == "Shadow"]
## t = -0.3112, df = 20.111, p-value = 0.6206
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -3.373306       Inf
## sample estimates:
## mean of x mean of y
##  18.54545  19.06122

#t test with alternative not equal
t.test(daysbelow32[ghog=="NOShadow"],daysbelow32[ghog=="Shadow"],
       alternative=c("two.sided"))

##
##  Welch Two Sample t-test
##
## data:  daysbelow32[ghog == "NOShadow"] and daysbelow32[ghog == "Shadow"]
## t = -0.3112, df = 20.111, p-value = 0.7588
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

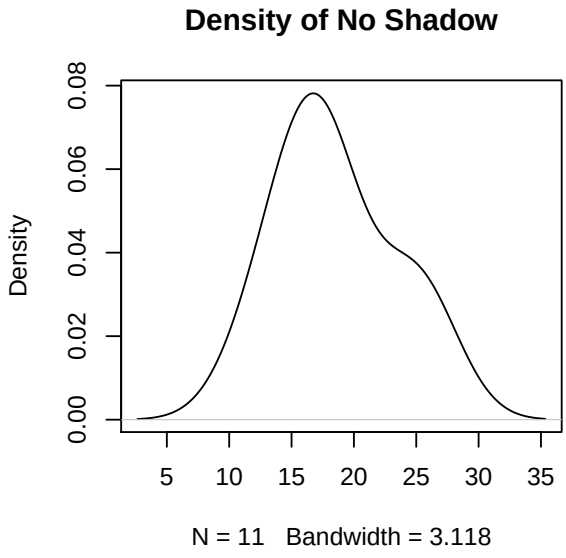
```
##  -3.971524  2.939984
## sample estimates:
## mean of x mean of y
##  18.54545  19.06122

pt(-.3112,20.11)

## [1] 0.3794258
```

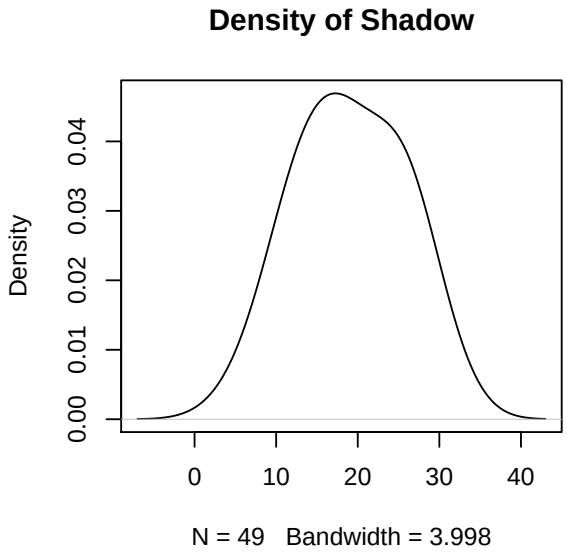
With a p-value of .62 for the alternative of greather than and a p-value of .76 for the alternative of equal our results imply that the difference in mean number of days below 32 when the groundhog does not see his shadow is statistically insignificant from the number of days below 32 degrees for the day he sees his shadow. A t-test is only valid if the data we have is approximately distributed normally. As such we generate histograms and smoothed density plots of the days below 32 degrees catagorized by whether the all knowing groundhog saw his shadow. We then create a pretty graph using ggplot2 and the wesanderson package. This step is crucial as everyone loves a pretty graph.

```
#Histograms to check normality
par(mfrow=c(2,2))
hist(daysbelow32[ghog=="NOShadow"],nclass=8,main="Historgram of No Shadow")
hist(daysbelow32[ghog=="Shadow"],nclass=8,main="Historgram of Shadow")

#Density plot to check normality
plot(density(daysbelow32[ghog=="NOShadow"],adjust=1.5),main="Density of No Shadow")
plot(density(daysbelow32[ghog=="Shadow"],adjust=1.5),main="Density of Shadow")
```



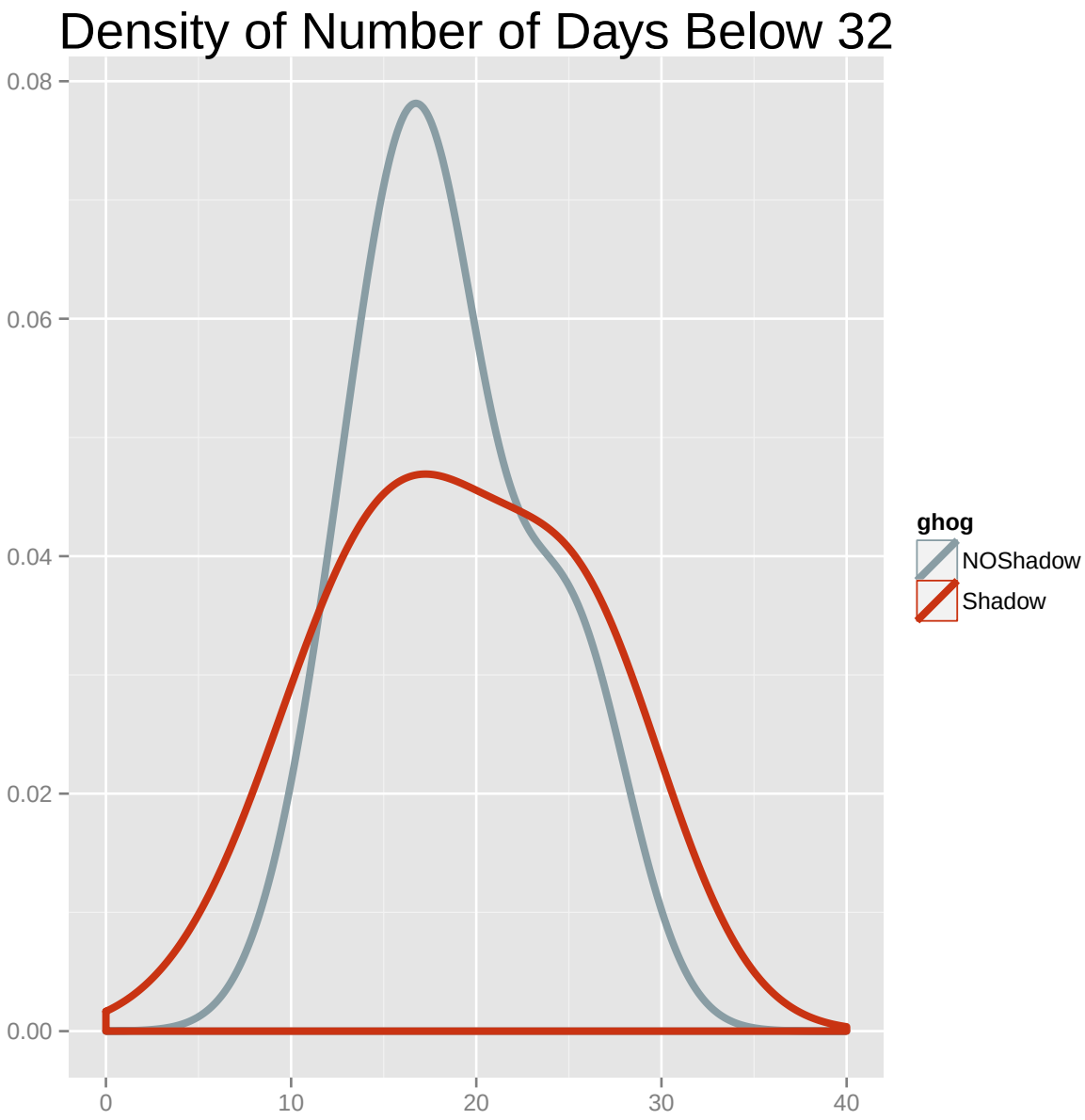







```
# Pretty ggplot to check normality

library(ggplot2)
library(wesanderson)

pretty <- ggplot(Ghog2,aes(colour=ghog))+
  scale_color_manual(values = wes.palette(2, "Royal1"))+
  geom_density(aes(x=daysbelow32),adjust=1.5,,size=1.5 )  +
  ggtitle("Density of Number of Days Below 32")+
    theme(axis.title.x = element_blank(), axis.title.y = element_blank())+
   theme(plot.title = element_text(size = rel(1.8))) + xlim(0,40)
```

```
pretty
```



In conclusion, with our plots showing an approximately normal distribution for our rather small samples, we conclude the t-test is valid. Our results imply that the difference in mean number of days below 32 degrees when the groundhog does not see his shadow is statistically insignificant from the number of days below 32 degrees for the day he sees his shadow.