

Alex_Williams_A1

Alexander Williams

2024-09-17

2.

a)

I chose to add each datapoint one at a time to hopefully reduce errors when inputting values

```
plants<-data.frame(Growth = c(14.6),
                    Plot    = c(1),
                    Pot     = c(1),
                    Treatment = c(1))

# plot 1, treatment 1
plants<-rbind(plants, list(15.2, 1, 1, 1))
plants<-rbind(plants, list(13.2, 1, 2, 1))
plants<-rbind(plants, list(12.9, 1, 2, 1))
plants<-rbind(plants, list(16.4, 1, 3, 1))
plants<-rbind(plants, list(12.2, 1, 3, 1))

# plot 2, treatment 1
plants<-rbind(plants, list(18.5, 2, 1, 1))
plants<-rbind(plants, list(16.7, 2, 1, 1))
plants<-rbind(plants, list(22.2, 2, 2, 1))
plants<-rbind(plants, list(18.8, 2, 2, 1))
plants<-rbind(plants, list(24.7, 2, 3, 1))
plants<-rbind(plants, list(20.3, 2, 3, 1))

# plot 1, treatment 2
plants<-rbind(plants, list(7.1, 1, 1, 2))
plants<-rbind(plants, list(7.7, 1, 1, 2))
plants<-rbind(plants, list(6.8, 1, 2, 2))
plants<-rbind(plants, list(6.0, 1, 2, 2))
plants<-rbind(plants, list(10.0, 1, 3, 2))
plants<-rbind(plants, list(8.3, 1, 3, 2))

# plot 2, treatment 2
plants<-rbind(plants, list(9.7, 2, 1, 2))
plants<-rbind(plants, list(8.8, 2, 1, 2))
plants<-rbind(plants, list(6.8, 2, 2, 2))
plants<-rbind(plants, list(9.0, 2, 2, 2))
plants<-rbind(plants, list(10.4, 2, 3, 2))
plants<-rbind(plants, list(11.3, 2, 3, 2))

plants
```

```
##      Growth Plot Pot Treatment
```

## 1	14.6	1	1	1
## 2	15.2	1	1	1
## 3	13.2	1	2	1
## 4	12.9	1	2	1
## 5	16.4	1	3	1
## 6	12.2	1	3	1
## 7	18.5	2	1	1
## 8	16.7	2	1	1
## 9	22.2	2	2	1
## 10	18.8	2	2	1
## 11	24.7	2	3	1
## 12	20.3	2	3	1
## 13	7.1	1	1	2
## 14	7.7	1	1	2
## 15	6.8	1	2	2
## 16	6.0	1	2	2
## 17	10.0	1	3	2
## 18	8.3	1	3	2
## 19	9.7	2	1	2
## 20	8.8	2	1	2
## 21	6.8	2	2	2
## 22	9.0	2	2	2
## 23	10.4	2	3	2
## 24	11.3	2	3	2

b)

```
plants[order(plants$Growth),]
```

##	Growth	Plot	Pot	Treatment
## 16	6.0	1	2	2
## 15	6.8	1	2	2
## 21	6.8	2	2	2
## 13	7.1	1	1	2
## 14	7.7	1	1	2
## 18	8.3	1	3	2
## 20	8.8	2	1	2
## 22	9.0	2	2	2
## 19	9.7	2	1	2
## 17	10.0	1	3	2
## 23	10.4	2	3	2
## 24	11.3	2	3	2
## 6	12.2	1	3	1
## 4	12.9	1	2	1
## 3	13.2	1	2	1
## 1	14.6	1	1	1
## 2	15.2	1	1	1
## 5	16.4	1	3	1
## 8	16.7	2	1	1
## 7	18.5	2	1	1
## 10	18.8	2	2	1
## 12	20.3	2	3	1
## 9	22.2	2	2	1
## 11	24.7	2	3	1

c)

```
attach(plants)
mean <- sum(Growth) / length(Growth)
mean
```

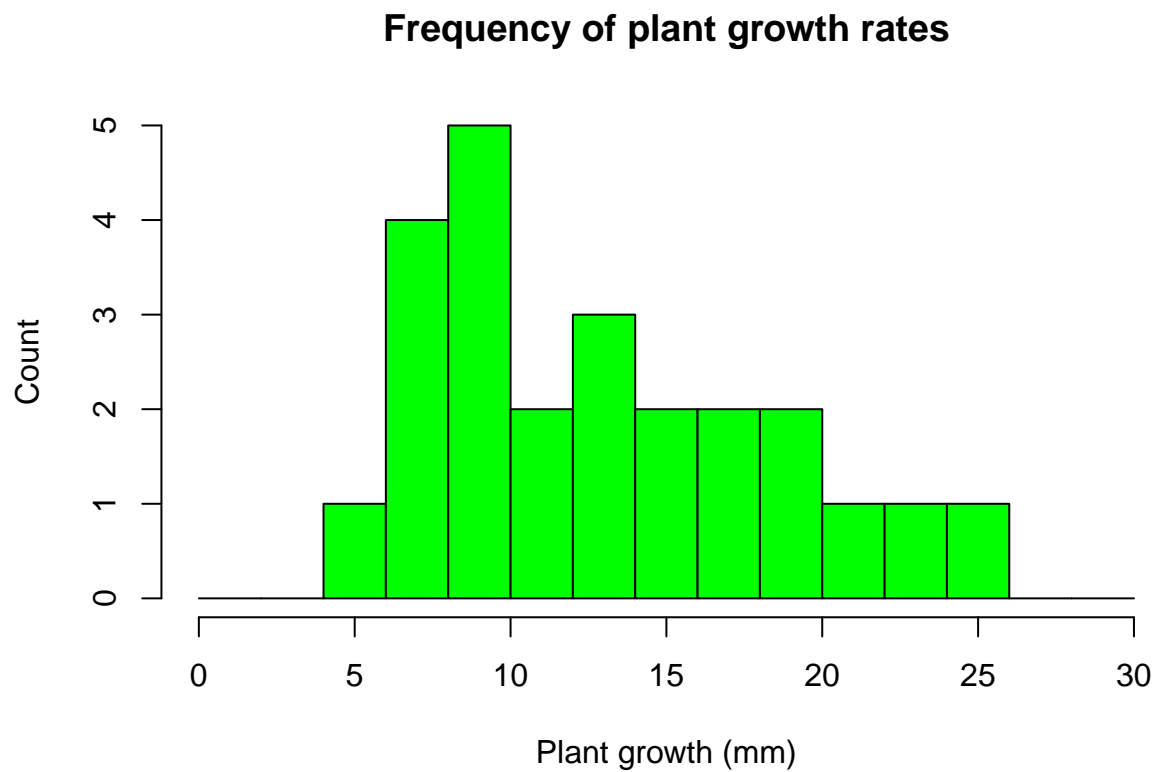
```
## [1] 12.81667
```

```
sd<- sd(Growth)
sd
```

```
## [1] 5.296813
```

d)

```
hist(Growth,
     breaks = seq(from=0, to=30, by=2),
     main = "Frequency of plant growth rates",
     xlab = "Plant growth (mm)",
     ylab = "Count",
     col = 'green')
```



```
detach(plants)
```

3.

```
get.variance<-function(y)
{
  mean = sum(y) / length(y)
  sample.var<- (sum(y^2) - (length(y) * mean^2)) / (length(y) - 1)
  sample.var
}
y<-c(11,11,10,8,11,3,15,11,7,6)
get.variance(y)
```

```
## [1] 11.34444
```

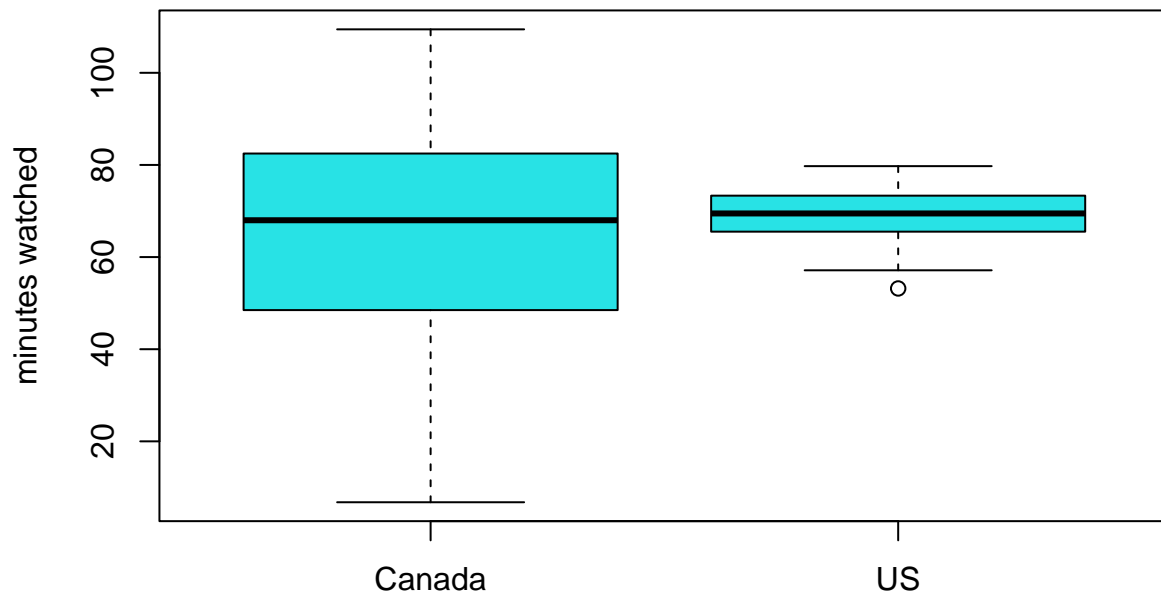
4.

a)

```
tv<-read.table(file="~/Documents/STAT 359/data/tv.txt",
               sep="",
               header=TRUE)
attach(tv)
Canada<-Canada[!is.na(Canada)]
US<-US[!is.na(US)]

boxplot(Canada, US,
        names=c("Canada", "US"),
        ylab="minutes watched",
        main="TV Watching Time",
        col=5)
```

TV Watching Time



```
summary(tv)
```

```
##      Canada      US
## Min.   : 6.781   Min.   :53.19
## 1st Qu.: 48.667   1st Qu.:65.58
## Median : 67.995   Median :69.47
## Mean   : 64.313   Mean   :69.33
## 3rd Qu.: 82.340   3rd Qu.:73.21
## Max.   :109.433   Max.   :79.74
## NA's   :10
```

While the Canada watch time and the US watch time have similar means and medians, US watch time seems to have a much tighter distribution.

c)

```
z.test<-function(y1, y2, H1)
{
  est<- mean(y1) - mean(y2)
  param <- 0 # since H0: mu1 - mu2 = 0
  t.obs <- (est - param) / (sqrt((var(y1) / length(y1)) + (var(y2) / length(y2))))
  if(H1=='less')
  {
    return.val <- pnorm(t.obs)
  }
  else if(H1 == 'greater')
  {

```

```

    return.val <- 1-pnorm(t.obs)
  }
  else if (H1 == 'two.sided')
  {
    return.val <- 2 * (1 - pnorm(abs(t.obs)))
  }
  return.val
}

```

c)

```

pval.less <- z.test(Canada, US, 'less')
pval.less

```

```
## [1] 0.02208637
```

```

pval.greater <- z.test(Canada, US, 'greater')
pval.greater

```

```
## [1] 0.9779136
```

```

pval.ts <- z.test(Canada, US, 'two.sided')
pval.ts

```

```
## [1] 0.04417275
```

d)

Since the goal of the study was to see if Canadian teenagers watched less TV than American teenagers, so we use the alternative hypothesis $H_1 : \mu_{Canada} - \mu_{US} < 0$, calculated with `z.test(Canada, US, 'less')`. This returns a p-value of ~0.0221. This means that about 2.21% of the time, if we were to run the experiment again, the result would be at least as extreme as this one.

We did not declare an alpha value before doing the experiment, so let us choose a standard value of $\alpha = 0.05$. since the p-value is less than alpha, we have evidence to reject the null hypothesis that Canadian and American teenagers watch the same amount of TV. Or, in other words, we have reason to believe that Canadian teenagers watch less TV than American teens.