Appendix 1- Part I

In 2018 research was conducted to find the price of "X" seeds per metric ton. "X" is a plant that would be introduced in the US market in 2019. The research compiled information from all over the United States and gave a low, medium, and high price for metric ton of "X" seeds. The research was conducted to gather data concerning the cost related to growing this plant. Note: One should consider a significance level of 0.05.

In 2019, when "X" was introduced to the US market, the selling price was recorded for each state. In 2030, when "X" has been part of the economy for more than 10 years the selling price was recorded for each state. We want to compare these with the costs predicted found in research conducted before 2019. It established a high price for metric ton of "X" seeds of 19,841.62, a medium price metric ton of "X" seeds of 8,818.50, and lastly, a low price metric ton of "X" seeds 4,406.06.

See dataset used below:

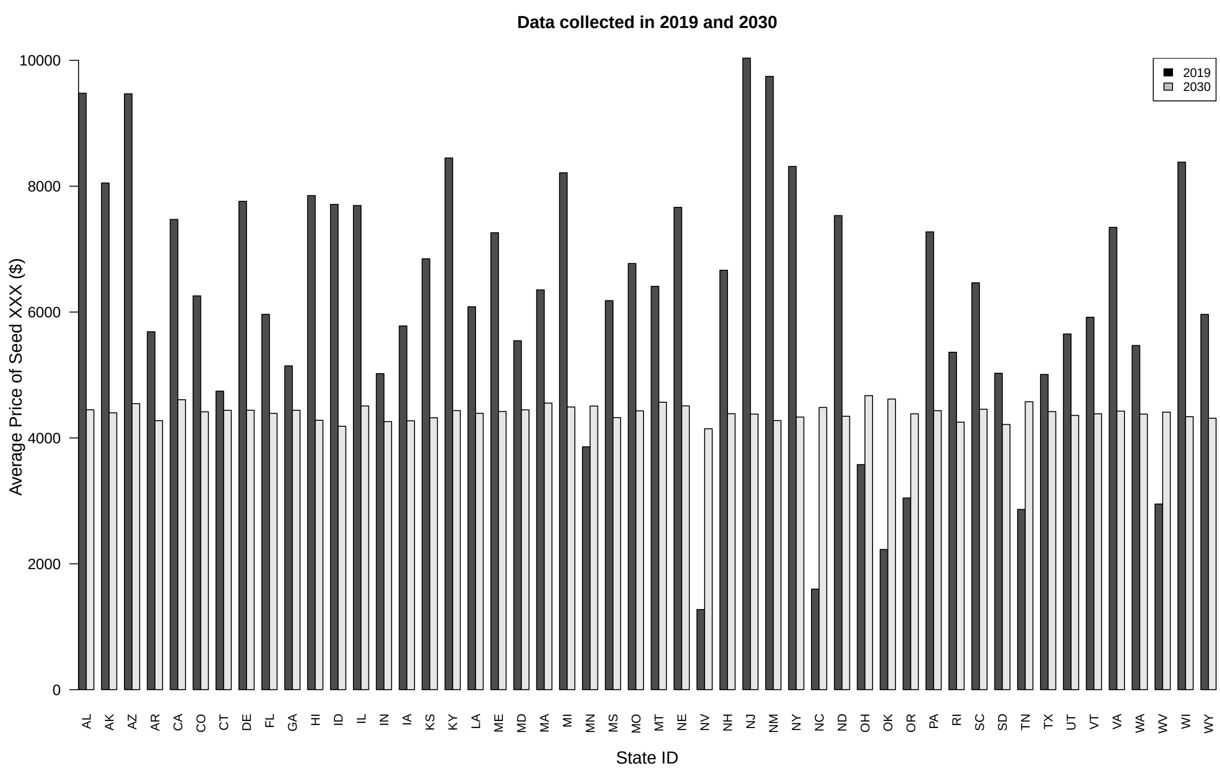
Table 1: Price_Data_2019_2030

State_ID	Price2019	Price2030
AL	9478	4447
AK	8051	4399
AZ	9468	4545
AR	5687	4275
CA	7472	4607
CO	6257	4415
CT	4744	4439
DE	7760	4440
FL	5964	4390
GA	5145	4440
HI	7851	4281
ID	7711	4186
IL	7693	4508
IN	5022	4260
IA	5780	4272
KS	6847	4320
KY	8448	4435
LA	6084	4391
ME	7261	4420
MD	5544	4446
MA	6352	4554
MI	8214	4492
MN	3859	4507
MS	6182	4323
MO	6772	4430
MT	6409	4568
NE	7664	4509
NV	1273	4145
NH	6665	4384
NJ	10035	4379
NM	9745	4277
NY	8314	4332
NC	1598	4485
ND	7533	4345
OH	3577	4672
OK	2228	4618

State_ID	Price2019	Price2030
OR	3046	4383
PA	7274	4433
RI	5362	4251
SC	6465	4456
SD	5028	4214
TN	2864	4575
TX	5009	4419
UT	5652	4358
VT	5918	4383
VA	7346	4427
WA	5469	4379
WV	2950	4410
WI	8382	4337
WY	5963	4313

Appendix 1.1 Data Visualization

```
dataplot<-rbind(Price2019,Price2030)
barplot(dataplot,beside=T, main="Data collected in 2019 and 2030", xlab="State ID",
ylab="Average Price of Seed XXX ($)", names.arg=Price_Data_2019_2030$State_ID,
cex.names=0.8, cex.lab= 1.2, las=2, xaxs = "i")
legend(x="topright",legend=c('2019','2030'), fill=c('black', 'gray'), cex = 0.8)
```



Appendix 1.2 *t*-test 2019 Data

Look into the data collected in 2019. Does the predicted data low, medium or high differ significantly from the data collected in 2019?

```r
#Price_Data_2019 is only the data collected for 2019.
Price_Data_2019<-Price_Data_2019_2030[,2]
Mu_High<-19841.62
Mu_Medium<-8818.50
Mu_Low<-4406.06
```

t-test High Price

The null hypothesis is that the mean of the high price for metric ton of "X" seeds (Mu_High) and mean of Price_Data_2019 (Mu_2019) are the same. The alternative hypothesis is that they are not the same.

Null hypothesis: Mu_High= Mu_2019

Alternative hypothesis: Mu_High!= Mu_2019

```r
t.test(Price_Data_2019, mu=Mu_High,conf.level=.95)
```

```
##
##  One Sample t-test
##
## data:  Price_Data_2019
## t = -46.995, df = 49, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 19841.62
## 95 percent confidence interval:
##  5646.175 6810.425
## sample estimates:
## mean of x
##    6228.3
```

Reject the Null hypothesis- The predicted high value and the average 2019 price are statistically significantly different considering the data collected in 2019, with a significance level of 0.05.

t-test Medium Price

The null hypothesis is that the mean of the medium price for metric ton of "X" seeds (Mu_Medium) and mean of Price_Data_2019 (Mu_2019) are the same. The alternative hypothesis is that they are not the same.

Null hypothesis: Mu_Medium= Mu_2019

Alternative hypothesis: Mu_Medium!= Mu_2019

```r
t.test(Price_Data_2019,mu=Mu_Medium,conf.level=.95)
```

```
##
##  One Sample t-test
##
## data:  Price_Data_2019
## t = -8.9417, df = 49, p-value = 7.202e-12
## alternative hypothesis: true mean is not equal to 8818.5
## 95 percent confidence interval:
```

```
##  5646.175 6810.425
## sample estimates:
## mean of x
##    6228.3
```

Reject the Null hypothesis- The predicted medium value and the average 2019 price are statistically significantly different considering the data collected in 2019, with a significance level of 0.05.

t-test Low Price

The null hypothesis is that the mean of the low price for metric ton of "X" seeds (Mu_Low) and mean of Price_Data_2019 (Mu_2019) are the same. The alternative hypothesis is that they are not the same.

Null hypothesis: Mu_Low= Mu_2019

Alternative hypothesis: Mu_Low!= Mu_2019

```
t.test(Price_Data_2019,mu=Mu_Low,conf.level=.95)
```

```
##
##  One Sample t-test
##
## data:  Price_Data_2019
## t = 6.2906, df = 49, p-value = 8.33e-08
## alternative hypothesis: true mean is not equal to 4406.06
## 95 percent confidence interval:
##  5646.175 6810.425
## sample estimates:
## mean of x
##    6228.3
```

Reject the Null hypothesis- The predicted low value and the average 2019 price are statistically significantly different considering the data collected in 2019, with a significance level of 0.05.

Appendix 1.3 *t*-test 2030 Data

Now let's take a look at the data found in 2030.Does the predicted data low, medium or high differ significantly from the data collected in 2019?

```
#Price_Data_2030 is only the data collected for 2030.
Price_Data_2030<-Price_Data_2019_2030[,3]
Mu_High<-19841.62
Mu_Medium<-8818.50
Mu_Low<-4406.06
```

t-test High Price

The null hypothesis is that the mean of the high price for metric ton of "X" seeds (Mu_High) and mean of Price_Data_2030 (Mu_2030) are the same. The alternative hypothesis is that they are not the same.

Null hypothesis: Mu_High= Mu_2030

Alternative hypothesis: Mu_High!= Mu_2030

```r
t.test(Price_Data_2030, mu=Mu_High,conf.level=.95)
```

```
##
##  One Sample t-test
##
## data:  Price_Data_2030
## t = -971.89, df = 49, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 19841.62
## 95 percent confidence interval:
##  4373.563 4437.397
## sample estimates:
## mean of x
##   4405.48
```

Reject the Null hypothesis- The predicted high value and the average 2030 price are statistically significantly different considering the data collected in 2030, with a significance level of 0.05

t-test Medium Price

The null hypothesis is that the mean of the medium price for metric ton of "X" seeds (Mu_Medium) and mean of Price_Data_2030 (Mu_2030) are the same. The alternative hypothesis is that they are not the same.

Null hypothesis: Mu_Medium= Mu_2030

Alternative hypothesis: Mu_Medium!= Mu_2030

```r
t.test(Price_Data_2030,mu=Mu_Medium,conf.level=.95)
```

```
##
##  One Sample t-test
##
## data:  Price_Data_2030
## t = -277.85, df = 49, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 8818.5
## 95 percent confidence interval:
##  4373.563 4437.397
## sample estimates:
## mean of x
##   4405.48
```

Reject the Null hypothesis- The predicted medium value and the average 2030 price are statistically significantly different considering the data collected in 2030, with a significance level of 0.05.

t-test Low Price

The null hypothesis is that the mean of the low price for metric ton of "X" seeds (Mu_Low) and mean of Price_Data_2030 (Mu_2030) are the same. The alternative hypothesis is that they are not the same.

Null hypothesis: Mu_Low= Mu_2030

Alternative hypothesis: Mu_Low!= Mu_2030

```r
t.test(Price_Data_2030,mu=Mu_Low,conf.level=.95)
```

```
##
##  One Sample t-test
##
## data:  Price_Data_2030
## t = -0.036518, df = 49, p-value = 0.971
## alternative hypothesis: true mean is not equal to 4406.06
## 95 percent confidence interval:
##  4373.563 4437.397
## sample estimates:
## mean of x
##   4405.48
```

Fail to reject the Null hypothesis- The predicted low value and the average 2030 price are not statistically significantly different considering the data collected in 2030, with a significance level of 0.05.