Presidency University
Data Analysis Project

Suryadeep Ghosh

January 17, 2020

Introduction:

This project is basically about comparing two different types of drugs in case of decreasing high blood pressure. The two types of drugs considered here are (i) **Ayurved** and (ii) **Alopathy**.We are also asked to see that if ayurved and alopath works at all in case of hypertension or not.

Only one aspect of blood pressure is considered here, viz. **systolic standing**. High blood pressure is also known as **Hypertension**. A normal systolic blood pressure is **below 120**.

Risk factors causing hypertensions are:-

1. **Obesity**.
2. **Lack of Physical Activity**.
3. **Sodium Intake**.
4. **Alcohol and drug use,Caffeine intake**.

In these cases, we can change the risk factors by stopping doing them. But in some cases we cannot change the risk factors,

1. **Genetics**.
2. **Age**.
3. **Race**.
4. **Other deceases and conditions**.

So in these cases we use drugs to control blood pressure.

Again,the authors of a 2012 study note that diabetes and hypertension often occur together and may share some common causes.

These include:

1. **obesity**.
2. **inflammation**.

3. **oxidative stress**.
4. **insulin resistance**.

So often we see that diabetes and hypertension occurring at the same time.

Exploratory Data Analysis:

First we read the data file.After reading we use the 'head' function to see that there are **7 variables** in the dataset.

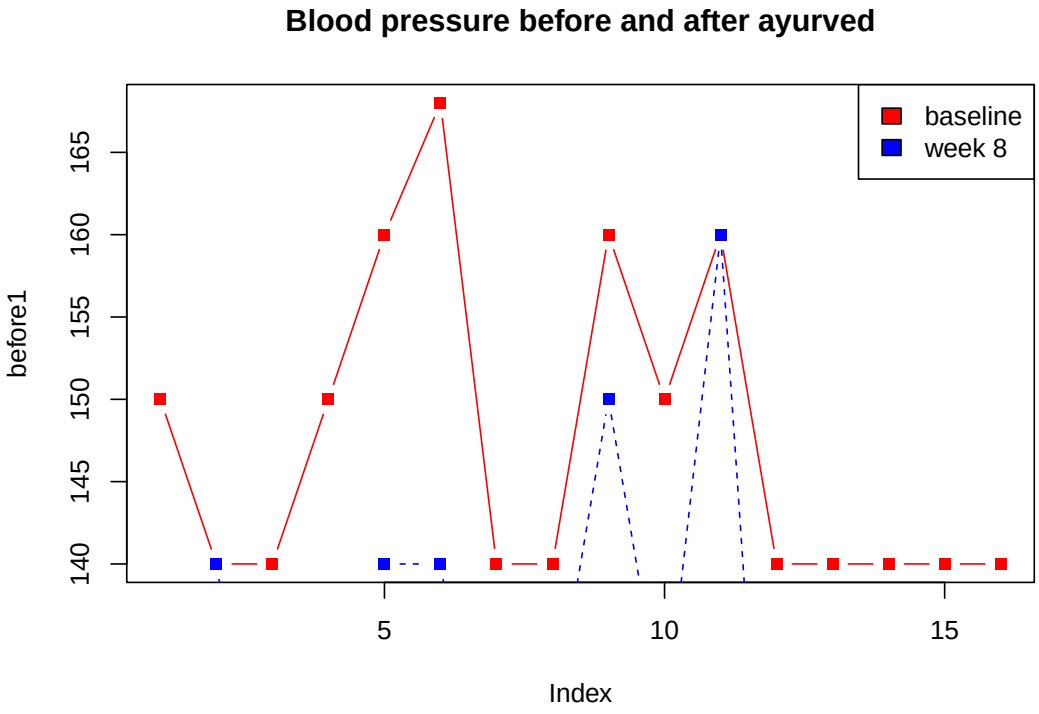
1. **"Age"** , which is numeric.
2. **"Sex"** , which is factor.
3. **"Duration.of.hypertension..yrs."**, which is numeric.
4. **"Duration.of.diabetes..yrs."** , which is numeric.
5. **"Systolic.BP..standing...mmHg.Baseline"** , which is numeric.
6. **"Systolic.BP..standing...mmHg.Week8"** , which is numeric.
7. **"Drug"** , which is factor.

- First we use the **'summary'** function to calculate the **summary measures** of the variables of the dataset such as mean,median,mode etc.

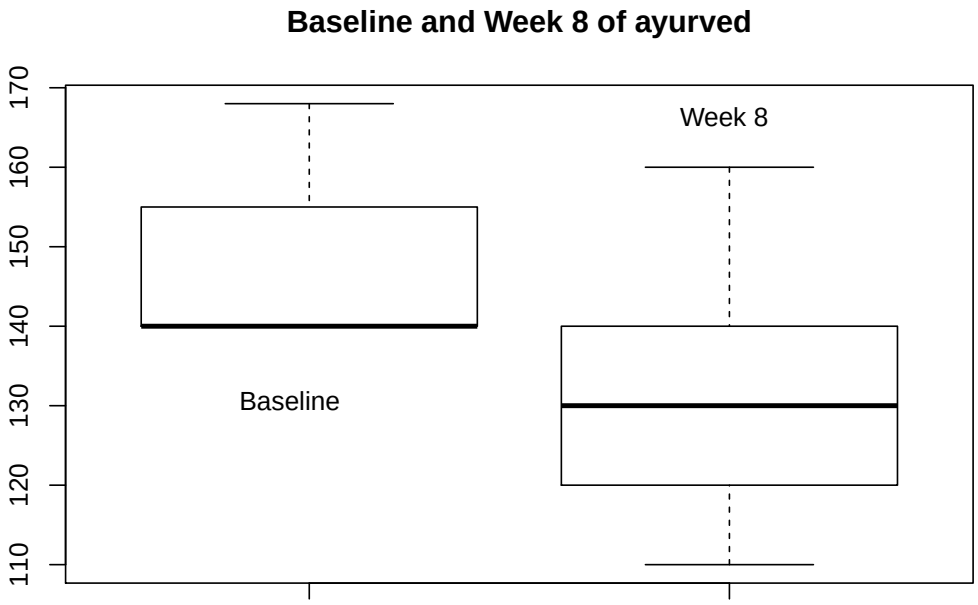
```
    Pat.no           Age        Sex    Duration.of.hypertension..yrs.
Min.   : 1.00   Min.   :32.00    : 0   Min.   : 0.170
1st Qu.:11.00   1st Qu.:49.00   F:17   1st Qu.: 0.670
Median :21.00   Median :53.00   M:16   Median : 2.000
Mean   :20.24   Mean   :52.82          Mean   : 3.619
3rd Qu.:29.00   3rd Qu.:60.00          3rd Qu.: 6.000
Max.   :40.00   Max.   :65.00          Max.   :13.000
Duration.of.diabetes..yrs. Systolic.BP..standing...mmHg.Baseline
Min.   : 0.170             Min.   :140.0
1st Qu.: 1.000             1st Qu.:140.0
Median : 2.000             Median :140.0
Mean   : 4.235             Mean   :147.5
3rd Qu.: 6.000             3rd Qu.:150.0
Max.   :15.000             Max.   :170.0
Systolic.BP..standing...mmHg.Week8 Drug
Min.   :110.0                        : 0
1st Qu.:120.0                       Al:17
Median :130.0                       Ay:16
Mean   :131.8
3rd Qu.:140.0
Max.   :170.0
```

- Then we use the **'cor'** function to calculate the **correlation** between the **time period of hypertension** and the **time period of diabetes**. The value of correlation we get is **0.09558932**,i.e there is some positive correlation between hypertension and diabetes.

- Then we need to see that **if ayurved is capable of curing hypertension**.So we first consider the readings of systolic blood pressure of those patients who took ayurvedic drug.A **line diagram** is done to see the trend of the data.

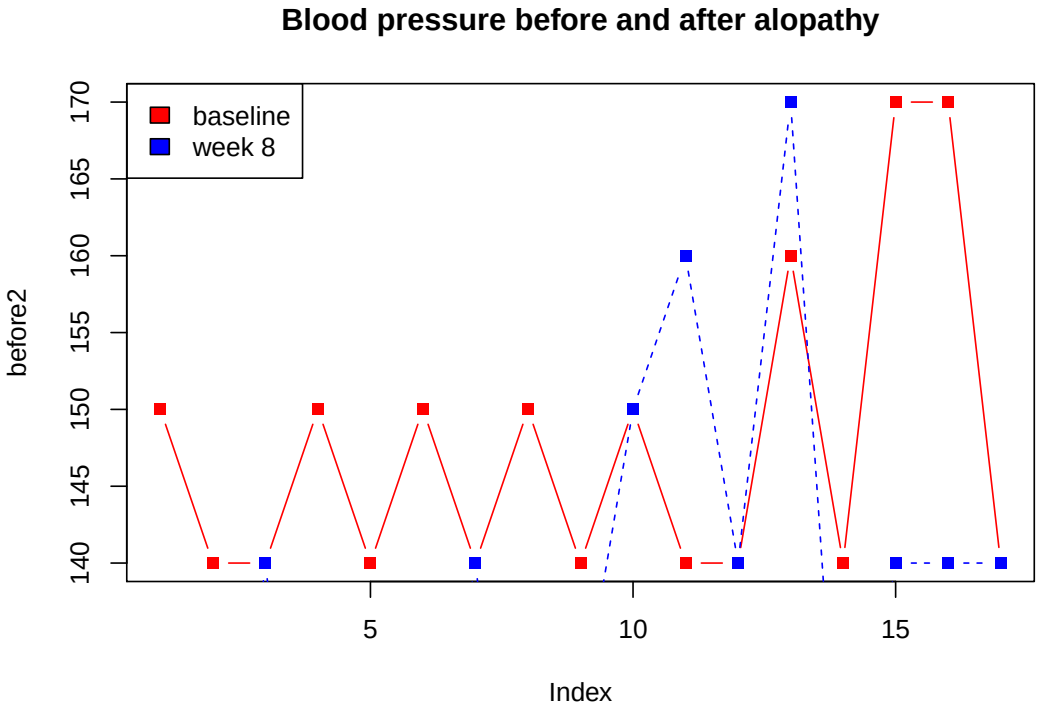


We plot the **boxplot** of readings taken at the baseline and on week 8 and compare them.We see that mean of readings of both cases **are not close at all** (mean of readings at baseline=147.375 and mean of readings on week 8=130.625).

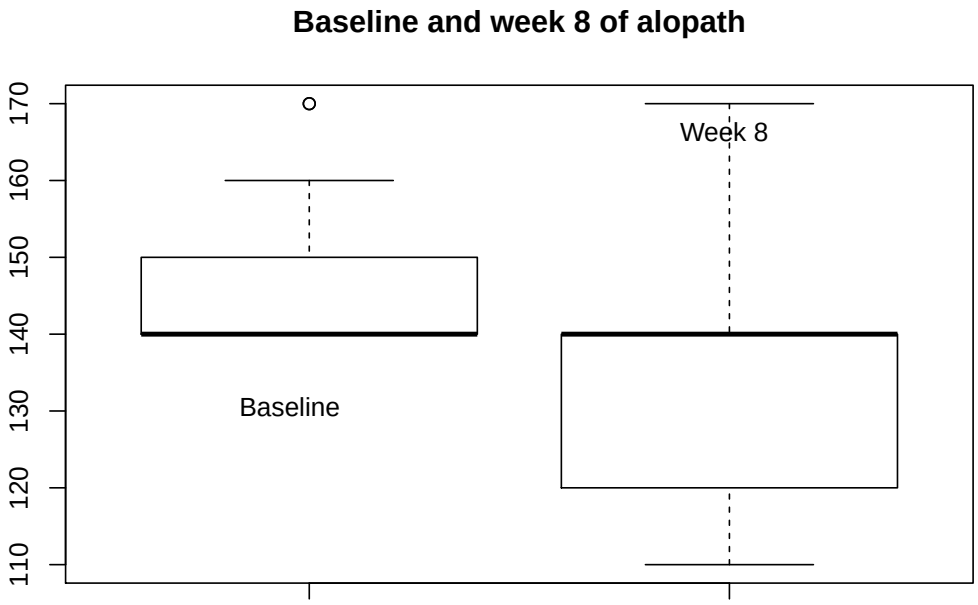


We leave it here for inferential data analysis.

- Then we try to check **if alopath is capable of curing hypertension**. So we first consider the readings of systolic blood pressure of those patients who took alopathic drug. A **line diagram** is done to see the trend of the data.



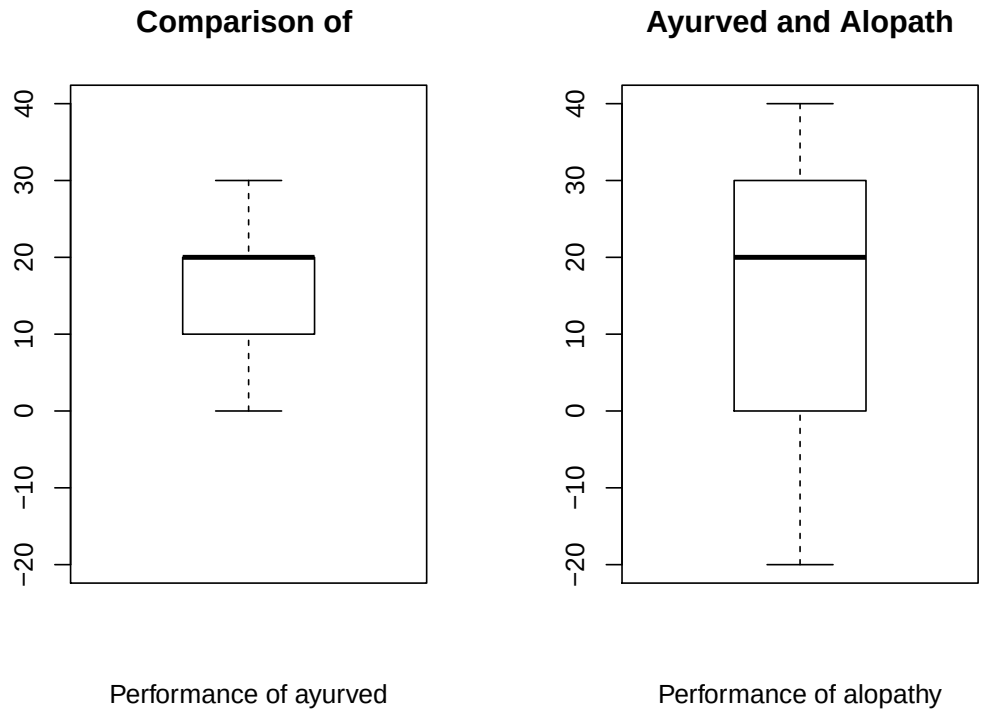
We plot the **boxplot** of readings taken at the baseline and on week 8 and compare them.



We see that the mean of readings of both cases **are not close at all** (mean of readings at baseline= 147.6471 and mean of readings on week 8=134.9412). We leave it for inferential data analysis.

- Then to check that **which of the two types of drugs is most effective** we take the difference of readings taken at baseline and readings taken on week 8 in both cases of ayurved and alopath.

We **boxplot** them to see that the **mean of the differences in case of ayrved and alopath are close** (mean for ayurved=16.7500 and mean for alopath=14.70588).



We leave it here for inferential data analysis.

Inferential Data Analysis:

Based on the exploratory data analysis now we have to do the inferential data analysis.

- **First**, we need to infer if there exists any correlation between The duration of hypertension and the duration of diabetes.Doing **t-test**(with level of significance 0.05) we see that the value the test statistic does not fall in the critical region. So we infer that the **null hypothesis is true** i.e **the value of ρ (correlation coefficient) is 0**. The 95% **confidence interval** is **[-0.3439573,0.3439573]**.

- **Then**, we need to infer about the difference of means of readings of baseline and readings of week 8 in case of Ayurved. Using The function **'t.test'** (with level of significance 0.05),we take the **null hypothesis to be that differences of two means is equal to 0** and **alternative hypothesis to be that the first mean is greater then second one**. After testing we see that the **p-value** of the test is **0.0001258** which is $<$ **0.05**. So we infer that **mean of readings of baseline is greater than mean of readings of week 8** i.e the **alternative hypothesis is true**.The 95% **confidence interval** is **[9.901547,∞)**.

- **Then,** we need to infer about the difference of means of readings of baseline and readings of week 8 in case of Alopath. Using The function **'t.test'** (with level of significance 0.05),we take the **null hypothesis to be that differences of two means is equal to 0** and **alternative hypothesis to**

be that the first mean is greater then second one. After testing we see that the **p-value** of the test is **0.003114** which is $<$ **0.05**. So we infer that **mean of readings of baseline is greater than mean of readings of week 8** i.e the **alternative hypothesis is true**.The 95% **confidence interval** is **[6.199885,∞)**.

- **Then,** we need to infer about the difference of means of differences of reading taken at baseline and on week 8 in case of both Ayurved and Alopath.Using the **'t.test'** (with level of significance 0.05), we take the **null hypothesis to be that difference of two means is equal to 0** and **alternative hypothesis to be that they are not equal**. After testing we see that the **p-value** of the test is **0.6885** which is $>$ **0.05**. So we infer that **the value of means are same** i.e the **null hypothesis is true**.The 95% **confidence interval** is **[-8.257207,12.345442]**.

Conclusion and Discussion:

Based on the both exploratory and inferential analysis,

1. **We see in the light of given data that diabetes does not cause hypertension as there is no statistically significant correlation among duration of diabetes and duration of hypertension.**
2. **In the light of given data we conclude that Ayurvedic drug cures hypertension as mean of readings taken before using the drug is significantly larger than mean of readings taken after using the drug for 8 weeks.**
3. **In the light of given data we conclude that Alopathic drug also cures hypertension as mean of readings taken before using the drug is significantly larger than mean of readings taken after using the drug for 8 weeks.**
4. **In the light of given data we conclude that both Ayurvedic and Alopathic drug work equally good in case of curing hypertension as difference of means of differences of reading taken at baseline and on week 8 in caes of both Ayurved and Alopath is statistically insignificant.**

So we conclude that **Ayurvedic and Alopathic drugs both can be used to cure hypertension**.

Appendix:

R codes:

```
B=read.table("C:\\Users\\LENOVO\\OneDrive\\Desktop\\BPSYScsv.csv",header=T,sep=",")
bp=na.omit(B)
head(bp)
summary(bp)
ay=subset(bp,Drug=="Ay")
al=subset(bp,Drug=="Al")
bl=bp$Duration.of.hypertension..yrs.
db=bp$Duration.of.diabetes..yrs.
cor(bl,db)
t=(r*sqrt(33-2))/(sqrt(1-r^2))
(t<=qt(0.025,31))&&(t>=qt(0.975,31))
u=qt(0.975,31)
a=(u^2)/31
sqrt(a/(a+1))
before1=ay$Systolic.BP..standing...mmHg.Baseline
after1=ay$Systolic.BP..standing...mmHg.Week8
plot(before1,type="b",pch=15,col="red",main="Blood pressure before and after ayurved
lines(after1,type="b",pch=15,col="blue",lty=2)
legend("topright",c("baseline","week 8"),fill=c("red","blue"))
before2=al$Systolic.BP..standing...mmHg.Baseline
after2=al$Systolic.BP..standing...mmHg.Week8
plot(before2,type="b",pch=15,col="red",main="Blood pressure before and after alopath
lines(after2,type="b",pch=15,col="blue",lty=2)
legend("topleft",c("baseline","week 8"),fill=c("red","blue"))
boxplot(before1,after1,xlab="Baseline and Week 8 of ayurved")
text(0.95366,130.5765,expression("Baseline"))
text(1.987617,166.2833,expression("Week 8"))
boxplot(before2,after2,xlab="Baseline and week 8 of alopath")
text(0.95366,130.5765,expression("Baseline"))
text(1.987617,166.2833,expression("Week 8"))
d1=(ay$Systolic.BP..standing...mmHg.Baseline-ay$Systolic.BP..standing...mmHg.Week8)
d2=(al$Systolic.BP..standing...mmHg.Baseline-al$Systolic.BP..standing...mmHg.Week8)
par(mfrow=c(1,2))
boxplot(d1,xlab="Performance of ayurved",ylim=c(-20,40))
boxplot(d2,xlab="Performance of alopathy",ylim=c(-20,40))
t.test(before1,after1,var.equal=TRUE,alt="greater")
t.test(before2,after2,var.equal=TRUE,alt="greater")
t.test(d1,d2,var.equal=TRUE)
```

Bibliography:

To get some facts about boxplot,hypothesis testing and hypertension the following references are followed:-

1. Fundamentals of Mathematical Statistics by S.C Gupta, V.K.Kapoor.
2. https://en.wikipedia.org/wiki/P-value.

3. http://chemocare.com/chemotherapy/side-effects/blood-pressure-changes.aspx.

Acknowledgment:

I would like to thank Atanu Kumar Ghosh sir for giving me this project and helping me learn about these interesting facts about hypertension and drugs used to cure it.I have learned a lot about topics related to this topic while doing this project.