Assignment01

8680643

Fei Yun

Question 1.

1-1

tapply(Elect15$LPV, Elect15$Prov, sum)

AB BC MB NB NL NS NT NU ON PE QC SK

473416 829816 268280 227764 165418 324816 9172 5619 2929393 51002 1515673 131681

YT

10887

1-2

> Elect<-subset(Elect15,Elect15$Prov=="BC")

> x1<-Elect$Tru

> w1<-Elect$Electors

> weighted.mean(x1,w1)

[1] 53.80233

This is vote for LIB and the leader Trudeau

1-3

> Elewin<-subset(Elect15,Elect15$LIB\_Win==1)

> sd(Elewin$ecc\_sat,na.rm=TRUE)

[1] 0.1280154

1-4

> quantile(sum(Elect15$Female,na.rm=TRUE),.28,na.rm=TRUE)

28%

16917

1-5

> mad(Elect$TO,center = median(Elect$TO),constant = 1.4826)

[1] 5.04084

Question 2.

2-1

A

> BQP<-aggregate(Elect15$BQP\_Win,by=list(Prov=Elect15$Prov),FUN=sum)

> CPC<-aggregate(Elect15$CPC\_Win,by=list(Prov=Elect15$Prov),FUN=sum)

> LIB<-aggregate(Elect15$LIB\_Win,by=list(Prov=Elect15$Prov),FUN=sum)

> NDP<-aggregate(Elect15$NDP\_Win,by=list(Prov=Elect15$Prov),FUN=sum)

> GRN<-aggregate(Elect15$GRN\_Win,by=list(Prov=Elect15$Prov),FUN=sum)

> Partwin<-cbind.data.frame(BQP$x,CPC$x,LIB$x,NDP$x,GRN$x)

> colnames(Partwin)<-c("BQP","CPC","LIB","NDP","GRN")

> rownames(Partwin)<-GRN$Prov

> Partwin<-Partwin/rowSums(Partwin)

> Partwin

BQP CPC LIB NDP GRN

AB 0.0000000 0.8529412 0.11764706 0.02941176 0.00000000

BC 0.0000000 0.2380952 0.40476190 0.33333333 0.02380952

MB 0.0000000 0.3571429 0.50000000 0.14285714 0.00000000

NB 0.0000000 0.0000000 1.00000000 0.00000000 0.00000000

NL 0.0000000 0.0000000 1.00000000 0.00000000 0.00000000

NS 0.0000000 0.0000000 1.00000000 0.00000000 0.00000000

NT 0.0000000 0.0000000 1.00000000 0.00000000 0.00000000

NU 0.0000000 0.0000000 1.00000000 0.00000000 0.00000000

ON 0.0000000 0.2727273 0.66115702 0.06611570 0.00000000

PE 0.0000000 0.0000000 1.00000000 0.00000000 0.00000000

QC 0.1282051 0.1538462 0.51282051 0.20512821 0.00000000

SK 0.0000000 0.7142857 0.07142857 0.21428571 0.00000000

YT 0.0000000 0.0000000 1.00000000 0.00000000 0.00000000

B

In NB,NL,NS,NT,NU,PE,YT the party LIB wins 100% is the highest.

2-2

A

GenderE<-Elect15[!(Elect15$Prov=="NU"|Elect15$Prov=="NT"|Elect15$Prov=="YT"),]

GenderE <- aggregate(GenderE[c("Male", "Female")], by = list(GenderE$Win\_Party), FUN = sum, rm.na = TRUE)

barchart(Group.1~ Male+Female,

data = GenderE,

beside = TRUE,main="Gender win by party",auto.key=list(),xlab="Population")

B

The Party CPC has higher female voters than male.

2-3Chart, bar chart

Description automatically generated

A

TruHist<-Elect15$Tru

hist(TruHist,main = "Trudeau supoorter saitisfaction")

Chart, histogram

Description automatically generated

B

The range of 50-55 has highest frequency

2-4

GenderE<-Elect15[!(Elect15$Prov=="NU"|Elect15$Prov=="NT"|Elect15$Prov=="YT"),]

bwplot(Lib~Win\_Party,data=GenderE,main="Platical Party Saitisfaction by Party",ylab="Saitisfaction" ,pch="|")

Chart, box and whisker chart

Description automatically generated

B

In LIB party, the party LIB has most support,the highest and lowest saitisfaction is high and median satisfaction is the highest in the 5.

C

In BQP party, the party LIB has lest support,because the median support is the lowest and major people saitisfacation is low.

D

In CPC party, the party has biggest variability. It has 35 range more than LIB’s range of 25 and NDP’s range of 25

2-5

A

TruHist<-Elect15$Tru

hist(TruHist,main = "Trudeau supoorter saitisfaction")

Chart, histogram

Description automatically generated

B

LibHist<-Elect15$Lib

hist(LibHist,main = "LIB supoorter saitisfaction")

Chart, histogram

Description automatically generated

C

plot(TruHist,LibHist,main = "PartyLeader vs Party")

Chart, scatter chart

Description automatically generated

D

abline(lm(LibHist~TruHist))Chart, scatter chart

Description automatically generated

According to the graph, we can know Trudeau supports and LIB Party supports have positive correlation. People who support Trudeau mostly support LIB Party as well.

E

cor.test(TruHist,LibHist)

Pearson's product-moment correlation

data: TruHist and LibHist

t = 58.373, df = 334, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.9436836 0.9629878

sample estimates:

cor

0.9543213

Question 3.

3-1

A

qqnorm(Elect15$LPV)

qqline(Elect15$LPV)

Chart, line chart, histogram

Description automatically generated

B

t.test(Elect15$LPV)

One Sample t-test

data: Elect15$LPV

t = 50.092, df = 337, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

19734.61 21347.86

sample estimates:

mean of x

20541.23

C

According to the T-test result, the LIB party Votes should be Binomial distribution

3-2

LIB\_Win<-mean(subset(Elect15$LPV,Elect15$LIB\_Win==1))

25089.4

LIB\_LOSE<-mean(subset(Elect15$LPV,Elect15$LIB\_Win==0))

15107.06

LIB\_Win<-subset(Elect15$LPV,Elect15$LIB\_Win==1)

LIB\_LOSE<-subset(Elect15$LPV,Elect15$LIB\_Win==0)

var.test(LIB\_Win,LIB\_LOSE,alternative="two.sided")

F test to compare two variances

data: LIB\_Win and LIB\_LOSE

F = 1.8082, num df = 183, denom df = 153, p-value = 0.0001744

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

1.330476 2.447126

sample estimates:

ratio of variances

1.808186

P value is less than 0.05, two group of data has great difference.

B

Because the variances of the two group is not equal, we can not use T-test here. We can see the result the ratio of variances is over 1.8 so the two data have great differences.

C

The ratio of variances is bigger than 1.