3174 SUHANA RAUTHAR

Are there Equal Opportunities in Science and Arts Stream

R CODE:

Setting up directory to import and view data:

```
getwd()
setwd("/cloud/project")
data<-read.csv("STATS.csv", header = TRUE, sep =",")
data
summary(data)
```

Descriptive statistics for numerical variable

```
income<-c(25000,30000,35000,20000,100000,40000,70000,15000,25000,12000,12000
,70000,30000,10000,20000,100000,20000,22000,14000,14000,10000,28000,15000,30
000,40000,20000,45000,40000,12000,100000,32000,12000,50000,400000,12000,2200
0,20000,35000,18000,30000,120000,25000,100000,30000,100000,50000,1000000000,
50000,250000,80000,200000,7500000,600000,50000,25000)
```

Measure of central tendency

```
mean(income)
median(income)
max(income)
min(income)
```

Measures of dispersion

```
range(income)
var(income)
sd(income)
summary(income)
```

T TEST

```
str(data)
  G=data$Gender
  I=data$Income
```

one sample t-test for Income variable

```
t.test(l)
```

TWO SAMPLE T TEST

1) **two sample t-test(Income~Gender)**

```
t.test(l~G)
```

2) **two sample t-test(Income ~ is there better opportunities in science)**

```
x=data$Better_opportunities_sciencegraduates
t.test(l~x)
```

NORMALITY TEST->SHAPIRO TEST FOR Numeric variable(Income)

```
shapiro.test(l)
```

CHI-SQUARE TEST

1) **(gender~stream)**

```
table<-c(5, 13, 22, 1, 8, 6)
m<-matrix(table, nrow=2, byrow =TRUE)
rownames(m)<-c("F", "M")
colnames(m)<-c("Arts", "Commerce", "Science")
print("Contingency Table:")
print(m)
chisq.test(m)
```

2) **chi-sq test for (switched_stream~stream)**

```
table_1<-c(5, 13, 21, 1, 8, 7)
m<-matrix(table_1, nrow=2, byrow=TRUE)
rownames(m)<-c("No", "Yes")
colnames(m)<-c("Arts", "Commerce", "Science")
print("Contingency Table:")
print(m)
chisq.test(m)
```

3) **chi-sq test for (willing to choose same stream again~stream)**

```
table_2<-c(1, 7, 6, 5, 14, 22)
m<-matrix(table_1, nrow=2, byrow=TRUE)
rownames(m)<-c("No", "Yes, Definitely")
colnames(m)<-c("Arts", "Commerce", "Science")
print("Contingency Table:")
print(m)
```

```
chisq.test(m)
```

ONE WAY ANOVA

1) **(factor_increasing_opportunities~Income)**

```
technical_knowledge<-c(25000,30000,20000,40000,70000,70000,30000,20000,20000,
14000,40000,45000,22000,20000,35000,18000,50000,1000000000,250000,50000,250
00)
confidence_extracurriculars<-c(35000,100000,15000,25000,12000,10000,100000,2200
0,14000,30000,20000,40000,12000,100000,32000,12000,50000,12000,120000,25000,
30000,80000,600000)
inherent_intelligence<-c(12000,10000,28000,15000,400000,30000,100000,100000,500
00,200000,7500000)
Combined_factors<-data.frame(cbind(technical_knowledge,
confidence_extracurriculars, inherent_intelligence))
Combined_factors
summary(Combined_factors)

Stacked_Groups<-stack(Combined_factors)
Stacked_Groups

Anova_results<-aov(values~ind, data = Stacked_Groups)
Anova_results
```

2) **one-way anova (stream~Individual perception)**

```
Arts<-c(4,3,2,5,5,5)
Commerce<-c(3,3,3,4,3,4,3,5,5,5,4,2,1,2,2,5,5,5,4,4,4)
Science<-c(5,5,4,3,4,3,2,2,4,4,3,2,3,3,3,3,3,2,2,3,4,3,2,3,4,5,3,3)
Combined_Streams<-data.frame(cbind(Arts, Commerce, Science))#combines the data
into a single data set
Combined_Streams #shows spreadsheet like results
summary(Combined_Streams) #min, median, mode, max

Stacked_Groups<-stack(Combined_Streams)
Stacked_Groups

Anova_results<-aov(values~ind, data = Stacked_Groups)
Anova_results
```

3) **one way anova(Stream~Income)**

```
Arts<-c(15000,10000,12000,30000,25000,50000)
Commerce<-c(30000,100000,25000,12000,12000,45000,40000,12000,32000,400000,12000,120000,100000,30000,100000,1000000000,250000,80000,200000,7500000,600000)
Science<-c(25000,35000,20000,40000,70000,70000,30000,20000,100000,20000,22000,14000,14000,10000,28000,15000,30000,40000,20000,100000,50000,22000,20000,35000,18000,50000,50000,25000)
Combined_Streams<-data.frame(cbind(Arts, Commerce, Science))
Combined_Streams
Stacked_Groups<-stack(Combined_Streams)
Stacked_Groups

Anova_results<-aov(values~ind, data = Stacked_Groups)
Anova_results
```

4) **<u>one way anova(Education~Income)</u>**

```
PostGraduation<-c(25000,35000,20000,40000,70000,25000,20000,22000,30000,40000,20000,12000,50000,50000,1000000000,50000,250000,50000,25000)
UnderGraduation<-c(30000,100000,12000,20000,10000,28000,30000,120000,100000,30000,200000,600000)
Graduation<-c(15000,70000,30000,10000,14000,14000,15000,45000,40000,100000,32000,400000,12000,22000,20000,35000,18000,25000,100000,80000,7500000)
Highschoolorbelow<-c(30000,100000,12000,12000,100000,20000,10000,28000,12000,30000,120000,100000,30000,200000,600000)
Combined_edu<-data.frame(cbind(PostGraduation, UnderGraduation, Graduation, Highschoolorbelow))
Combined_edu
Stacked_Groups<-stack(Combined_edu)
Stacked_Groups

Anova_results<-aov(values~ind, data = Stacked_Groups)
Anova_results
```

<u>TWO WAY ANOVA</u> (Income ~ Stream + Education)

```
data$Stream <-as.factor(data$Stream)
data$Education<-as.factor(data$Education)
anova<-aov(Income ~Stream + Education, data = data)
anova
```

3174 SUHANA RAUTHAR

Non Parametric tests

WILCOXON Signed Rank Test (Alternative of one sample t test)

wilcox.test(data$Income)
#pval 0.00....1<0.05 at 5% los Reject H0

Mann Whitney U Test/ Wilcoxon Sum Rank Test (Alt of Unpaired/ independent sample t test)

(Income~Gender)
wilcox.test(data$Income~data$Gender, paired=FALSE, exact=FALSE)
#PVAL 0.02<0.05 at 5% los ACCEPT H0

Kruskal wallis test (Alt of One way ANOVA)

1) kruskal.test(data$Income~data$factor_increasing_opportunities) #(Income ~factor_increased opportunity)

2) kruskal.test(data$Individual.Perception.of.Equal.Opportunities~data$Stream) #(perceptionofequality~Stream)

3) kruskal.test(data$Income~data$Stream) #(Income~Stream)

Visualization plots

BOXPLOT

boxplot(data$Education~data$Stream, xlab="Stream", ylab="Education", main="Education vs Stream")
boxplot(data$Individual.Perception.of.Equal.Opportunities~data$Stream, xlab="Stream", ylab="Individual Perception", main="Individual Perception vs Stream")

BARPLOT

plot(x=data$Education, y=data$Stream, xlab="Education", ylab="Stream", main="Education vs Stream")