**1)**

Dataset we are using is adult dataset. Variable or fields or attributes in the dataset are age, work class, demogweight, education, education.num, marital.status, occupation, relationship, race, sex, capital.gain, capital.loss, hours.per.week, native.country, income,part.

Among all the above attributes of adults dataset we took age, work class , demogweight attributes. Age represents number of people, work class represents the type of work people involved in like private, state-government and self-employed. Demogweight represents economic status of people. Part represents partition rate .

We took a subset of adults with three variables and performed two partition test.

R code:

Below code represents 1 through 5 values as written in R code. We partitioned it into two data sets one is training set and other is testing set.

Here training set: In general training set we take a sample of data for the point estimation.

Testing set: By testing set we perform validation. In this training acts as a reference and testing acts as a validation.

> adults$part<-runif(length(adults$income))

> adults$part<-runif(length(adults$income),min=0,max=1)

> adults[1:5,c(1,2,3,16)]

age workclass demogweight part

1 39 State-gov 77516 0.7834955

2 50 Self-emp-not-inc 83311 0.6228512

3 38 Private 215646 0.7660214

4 53 Private 234721 0.5476672

5 28 Private 338409 0.8470722

> training<-adults[adults$part<=0.70,]

> testing<-adults[adults$part>0.7,]

> training[1:5,c(1,2,3,16)]

age workclass demogweight part

2 50 Self-emp-not-inc 83311 0.6228512

4 53 Private 234721 0.5476672

6 37 Private 284582 0.5527014

7 49 Private 160187 0.4422584

8 52 Self-emp-not-inc 209642 0.2456715

> testing[1:5,c(1,2,3,16)]

age workclass demogweight part

1 39 State-gov 77516 0.7834955

3 38 Private 215646 0.7660214

5 28 Private 338409 0.8470722

11 37 Private 280464 0.7484999

12 30 State-gov 141297 0.7740671

**2)** **R code:**

View(adults)

> training<-adults[adults$part<=0.50,]

> testing<-adults[adults$part>0.50,]

t.test(training[,c(1)],testing[,c(1)],alternative = c("two.sided"),mu=0,paired=FALSE,conf.interval=0.95)

Welch Two Sample t-test

data:  training[, c(1)] and testing[, c(1)]

t = -0.5777, df = 24997, p-value = 0.5635

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 -0.4393870  0.2393417

sample estimates:

mean of x mean of y

 38.55693  38.65696

Here we took training and testing and partitioned sample test of adult to below 50% and testing set greater than 50 percent. Projected the training set on testing set for validation.

From the above code we can define p-value: probability value equal to 0.5635

Now we perform a hypothesis test with the help of P-value and significance (Alpha):

Generally a hypothesis test means we are going to assert a value which is always a prediction. This prediction differs from other generalized prediction as we are asserting a educated guess and the probability of occurring a prediction is always maintained as maximum.

There are actually two types of hypothesis:

1.Null Hypothesis (H-sub-0).

2.Alternative Hypothesis(H-sub-a).

If the evidence is not present about the hypothesis then we accept null hypothesis (H-sub-0)

Here p-value=0.5635

Alpha=1-confidence Interval

As we maintained confident interval 95% equals to 0.95. As long as we maintain the confidence interval maximum then the chance of occurring will be maximum. So, we took 0.95.

Alpha = 1-0.95

= 0.05.

Two case come into picture they are:

P -value> Alpha: If this is case we accept null hypothesis and reject alternative hypothesis.

P-value<= Alpha: We accept alternative hypothesis and reject null hypothesis.

0.5635>0.05

By above illustration we accept Null Hypothesis and reject alternative hypothesis.

There is no evidence to reject H-sub-0 at a significance of Alpha = 0.05. So, we reject alternative and accept null hypothesis.

As we took Mu value equal to zero which represents two means are equal.The population mean number of age of all adults for training and testing set is 38.55693 38.65696(from R code).